AMERICAN
STATISTICAL ASSOCIATION.

NEW SERIES, No. 90. JUNE, 1910.

DERIVATION OF THE UNITED STATES MORTALITY TABLE BY OSCULATORY INTERPOLATION.

REGISTRATION STATES, TWELFTH CENSUS.

BY JAMES W. GLOVER, *Professor of Mathematics and Insurance, University of Michigan.*

INTRODUCTION.

The object of this paper is to derive a representative or standard mortality table from the best available population and vital statistics furnished by the United States Census Bureau. With this end in view I have made a study of the statistics for (1) aggregate males, and (2) native white, native parents, males. This paper deals with the first group, and the data employed are confined to the following eleven registration states: Connecticut, District of Columbia, Indiana, Maine, Massachusetts, Michigan, New Hampshire, New Jersey, New York, Rhode Island, and Vermont. The following statements, quoted from the official census reports of recent date, serve to define the term *registration states.*

"These are (the registration states) all states . . . for which the registration of deaths under local laws and ordinances was found to be sufficiently accurate for use by the Census Office." Abstract of the Twelfth Census, p. 178.

"For the decennial mortality statistics of the Twelfth Census, the 'registration area' was fixed after the records of deaths obtained from registration sources had been compared with the mortality returns of the census enumerators." Mortality Statistics, 1900–1904, p. xiv.

"The 'registration area'—that is to say, the states having

laws the results of whose operation have been accepted as giving practically complete mortality returns, . . ." Mortality Statistics, 1905, p. 5.

" . . . the completeness of registration of deaths for the state as a whole was equal to that of the minimum accepted for the registration area (90 per cent). . . ." Mortality Statistics, 1906, p. 27.

" . . . it is not yet certainly known whether any registration area, and especially any state area, in the United States has succeeded in obtaining an actual registration of at least 90 per cent of all births that occur — the minimum standard for the admission of a state or city to the registration area for deaths." Mortality Statistics, 1906, p. 16.

"The *registration area* of the United States, which embraces those states and separate cities accepted as having approximately complete registration of deaths based upon the requirement of compulsory burial permits, . . ." Bulletin 104, Mortality Statistics, 1908, p. 7.

It may fairly be assumed from the preceding statements that the best available data, namely, that of the registration area, is believed to represent not less than 90 per cent of the actual deaths occurring within such area. For the eleven states above mentioned the following table exhibits in quinquennial age groups the population as of June 1, 1900, and the deaths during the calendar year 1900. The figures relate to the aggregate group of males.

TABLE I.—FUNDAMENTAL DATA, UNITED STATES CENSUS BUREAU.

AGGREGATE MALES, REGISTRATION STATES.

AGE GROUP.	POPULATION JUNE 1, 1900.	DEATHS, 1900.	AGE GROUP.	POPULATION JUNE 1, 1900.	DEATHS, 1900.
5–9	998,284	4,688	55–59	341,797	8,475
10–14	911,960	2,646	60–64	275,523	9,333
15–19	881,784	4,286	65–69	201,124	10,125
20–24	915,560	6,412	70–74	140,705	10,221
25–29	911,834	7,171	75–79	85,026	9,348
30–34	831,959	7,258	80–84	40,356	6,791
35–39	756,051	7,671	85–89	13,704	3,378
40–44	645,445	7,482	90–94	2,997	1,037
45–49	517,413	7,230	95–99	551	220
50–54	444,151	7,934	100 and over	130	54

The population data were drawn from Table 2, pp. 18–99, Twelfth Census, Vol. II, the totals being found by adding together for each quinquennial age group the corresponding aggregate male population of the eleven registration states. The deaths, in quinquennial age groups, were taken from Table 4, pp. 60 61, line 52, Mortality Statistics, 1900–1904. From these data the final mortality table was derived by the method of osculatory interpolation in the manner hereafter explained.

THEORY OF OSCULATORY INTERPOLATION.

I next pass to the consideration of the formulæ employed in the reduction of the above data. One of the chief difficulties met with in adjusting rough data by the usual interpolation formulæ is that at the points where the interpolation curves meet there are sudden breaks in the values of the first differences which do not accord with our preconceived notion of the continuity of the function under discussion. Various methods have been resorted to in joining these interpolation curves; probably the most satisfactory, so far as results are concerned, is the *welding* process by means of sine curve factors employed in constructing the Registrar-General's English [1] Life Table, No. 6. The method employed in this paper, however, proceeds along quite different lines. The fundamental idea is to join two successive interpolation curves in such manner that for a certain abscissa they shall have a common ordinate, gradient, and radius of curvature. This is accomplished by making their first and second derivatives equal, respectively, at the point of intersection determined by the abscissa and common ordinate. In the case of interpolation curves of higher degree it will appear that the third and higher derivatives may be equal. This method has been termed "Osculatory Interpolation" because the successive interpolation curves have a common osculating circle at their points of junction; it was first devised by Sprague,[2] and later developed and applied by Karup,[3] King,[4] Buchanan,[5] and

[1] For description of method see Supplement, Part I, p. xvii, 65th Annual Report of the Registrar-General.

[2] Thomas Bond Sprague, Explanation of a New Formula for Interpolation, Jour. Inst. Actuaries, Vol. 22, p. 270.

[3] Johannes Karup, On a New Mechanical Method of Graduation, Trans. Second International Actuarial Congress, p. 78.

[4] George King, On the Construction of Mortality Tables from Census Returns and Records of Deaths, Jour. Inst. Actuaries, Vol. 42, p. 225.

[5] James Buchanan, Osculatory Interpolation by Central Differences; with an Application to Life Table Construction, Jour. Inst. Actuaries, Vol. 42, p. 369.

others. In developing the subject of osculatory formulæ I have not followed any of the above writers, but have generalized the osculatory formula along a line suggested by Lidstone.[1]

The problem is to find an interpolation curve y_x which shall be a blend of two curves u_x and v_x, so that

$$y_x = (1 - c_x)u_x + c_x v_x, \tag{1}$$

where c_x is a rational algebraic function of x, and u_x and v_x are defined as follows:

$$u_x = u_{-n+(x+n)} = u_{-n} + {}_{x+n}C_1\Delta u_{-n} + {}_{x+n}C_2\Delta^2 u_{-n} + \cdots + {}_{x+n}C_{2n}\Delta^{2n}u_{-n}, \tag{2}$$

$$v_x = v_{-(n-1)+(x+n-1)} = u_{-(n-1)} + {}_{x+n-1}C_1\Delta u_{-(n-1)} + {}_{x+n-1}C_2\Delta^2 u_{-(n-1)} + \cdots + {}_{x+n-1}C_{2n}\Delta^{2n}u_{-(n-1)}, \tag{3}$$

where $${}_mC_r = \frac{m(m-1)\cdots(m-r+1)}{1\cdot 2\cdots r}.$$

But $$u_{k+1} = u_k + \Delta u_k,$$

and hence $$c\Delta^s u_{k+1} = c\Delta^s u_k + c\Delta^{s+1}u_k.$$

Also $${}_{x+n-1}C_{r-1} + {}_{x+n-1}C_r = {}_{x+n}C_r,$$

and therefore

$$v_x = u_{-n} + {}_{x+n}C_1\Delta u_{-n} + \cdots + {}_{x+n}C_{2n}\Delta^{2n}u_{-n} + {}_{x+n-1}C_{2n}\Delta^{2n+1}u_{-n}, \tag{4}$$

whence

$$v_x = u_x + {}_{x+n-1}C_{2n}\Delta^{2n+1}u_{-n}. \tag{5}$$

It thus appears that u_x and v_x are rational, integral functions of x of degree $2n$ which pass through the ends of the first and last $2n+1$ ordinates, u_{-n}, $u_{-(n-1)}$, $\cdots u_{-1}$, u_0, u_1, $\cdots u_{n-1}$, u_n, u_{n+1}, respectively. They thus intersect or have in common the $2n$ points (k, u_k), where $k = -(n-1), \cdots -1, 0, 1, \cdots n-1, n$. Since by (1) the curve y_x passes through any point common to u_x and v_x, it must also pass through the above $2n$ points.

Let

$$y_x = u_x + \psi_x, \tag{6}$$

where ψ_x is a rational integral function of x of degree not less than $2n+1$. We have also by (5) the relation

$$v_x - u_x = {}_{x+n-1}C_{2n} \cdot \Delta^{2n+1}u_{-n}. \tag{7}$$

At the origin, $x = 0$, we assume that y_x and u_x, together with their first n derivatives, are equal, respectively, also at the point

[1] George J. Lidstone, Alternative Demonstration of the Formula for Osculatory Interpolation, Jour. Inst. Actuaries, Vol. 42, p. 394.

$x=1$, that y_x and v_x, together with their first n derivatives, are equal, respectively. Hence

$$\begin{aligned} y_0 &= u_0, \\ y_0' &= u_0', \\ y_0'' &= u_0'', \\ &\cdots \\ y_0^{(n)} &= u_0^{(n)}. \end{aligned} \qquad (8) \qquad \begin{aligned} y_1 &= v_1, \\ y_1' &= v_1', \\ y_1'' &= v_1'', \\ &\cdots \\ y_1^{(n)} &= v_1^{(n)}. \end{aligned} \qquad (9)$$

The relations (6) and (8) show that $\psi_0=\psi_0'=\psi_0''=\cdots=\psi_0^{(n)}=0$, hence ψ_x contains the factor x^{n+1}, while (5), (6), and the first one in (9) show that $\psi_1=0$, hence ψ_x contains the factor $x-1$. The remaining equations in (9) lead to the following set of relations:

$$\begin{aligned} \psi_1' &= ({}_{x+n-1}C_{2n}')_{x=1}\cdot\Delta^{2n+1}u_{-n}, \\ \psi_1'' &= ({}_{x+n-1}C_{2n}'')_{x=1}\cdot\Delta^{2n+1}u_{-n}, \\ &\cdots\cdots\cdots \\ \psi_1^{(n)} &= ({}_{x+n-1}C_{2n}^{(n)})_{x=1}\cdot\Delta^{2n+1}u_{-n}. \end{aligned} \qquad (10)$$

But

$$\begin{aligned} \psi_x &= x^{n+1}(x-1)\phi_x\,\Delta^{2n+1}u_{-n}/\underline{|2\,n} \\ &= (x-1)\xi_x\,\Delta^{2n+1}u_{-n}/\underline{|2\,n}, \end{aligned} \qquad (11)$$

where ϕ_x is a rational integral function of x of degree not less than $n-1$. If ϕ_x is taken of degree $n-1$, we have

$$\phi_x = a_0x^{n-1}+a_1x^{n-2}+\cdots+a_{n-1}, \qquad (12)$$

$$\xi_x = a_0x^{2n}+a_1x^{2n-1}+\cdots+a_{n-1}x^{n+1}. \qquad (13)$$

Let $\psi_x=\omega_x\,\xi_x\,\Delta^{2n+1}u_{-n}/\underline{|2\,n}$,

where $\omega_x=x-1$, $\omega_x'=1$, $\omega_x''=\omega_x'''=\cdots=0$,

whence $\omega_1=0$, $\omega_1'=1$, $\omega_1''=\omega_1'''=\cdots=0$.

Then $\psi_x'=\omega_x\xi_x'+\omega_x'\xi_x$,

$\psi_x''=\omega_x\xi_x''+2\,\omega_x'\xi_x'+\omega_x''\xi_x$,

$\cdots\cdots\cdots\cdots$

$\psi_x^{(n)}=\omega_x\xi_x^{(n)}+n\omega_x'\xi_x^{(n-1)}+c_2\omega_x''\xi_x^{(n-2)}+\cdots+c_n\omega_x^{(n)}\xi_x$,

where c_2, c_3, $\cdots$ c_n, are constants. Therefore

$$\psi_1^{(k)} = k\xi_1^{(k-1)}\cdot\Delta^{2n+1}u_{-n}/\underline{|2\,n}. \qquad (14)$$

Similarly, since ${}_{x+n-1}C_{2n}\Delta^{2n+1}u_{-n}$

$$\begin{aligned} &= (x-1)[(x+n-1)\cdots x(x-2)\cdots(x-n)]\Delta^{2n+1}u_{-n}/\underline{|2\,n} \\ &= \omega_x\,\mu_x\,\Delta^{2n+1}u_{-n}/\underline{|2\,n}, \end{aligned}$$

where

$$\mu_x = (x+n-1)\cdots x(x-2)\cdots(x-n), \tag{15}$$

we find that

$$({}_{x+n-1}C_{2n}^{(k)})_{x=1}\cdot\Delta^{2n+1}u_{-n} = k\mu_1^{(k-1)}\Delta^{2n+1}u_{-n}/\underline{|2n}, \tag{16}$$

and comparing equations (10), (14), and (16), we have

$$\xi_1^{(k-1)} = \mu_1^{(k-1)} \quad k=1, 2, \cdots n, \tag{17}$$

which furnish the n linear equations determining the n constants $a_0, a_1, \cdots a_{n-1}$, of ϕ_x.

Having thus determined ψ_x we can find c_x by the relation

$$\psi_x = c_x(v_x - u_x), \tag{18}$$

which taken with (7) gives

$$\psi_x = c_x\cdot{}_{x+n-1}C_{2n}\Delta^{2n+1}u_{-n},$$

whence

$$c_x = \frac{\psi_x}{{}_{x+n-1}C_{2n}\Delta^{2n+1}u_{-n}} = \frac{x^{n+1}\phi_x}{\mu_x},$$

and therefore we have

$$c_x = \frac{x^n(a_0x^{n-1}+a_1x^{n-2}+\cdots+a_{n-1})}{(x+n-1)\cdots(x+1)(x-2)\cdots(x-n)}. \tag{19}$$

SPECIAL OSCULATORY FORMULÆ.

We shall now apply the preceding theory to obtain the osculatory formulæ in a few special cases. The simplest case is where $n=1$. Here we have four ordinates y_{-1}, y_0, y_1, y_2, and u_x is defined by the first three and v_x by the last three. From equations (6) and (11) we have

$$y_x = u_x + x^2(x-1)a_0\Delta^2y_{-1}/\underline{|2}, \tag{20}$$

where

$$u_x = y_{-1} + (x+1)\Delta y_{-1} + \frac{(x+1)x}{\underline{|2}}\cdot\Delta^2y_{-1},$$

and (17) furnishes the relation $a_0=1$, and (19) makes $c_x = x$, whence from (1) we have

$$y_x = (1-x)u_x + xv_x.$$

The expression for y_x given in (20) easily transforms into the following formula obtained by Karup:[1]

$$y_x = y_0 + x\Delta y_0 + \frac{x(x-1)}{\underline{|2}}\Delta^2y_{-1} + \frac{x^2(x-1)}{\underline{|2}}\Delta^3y_{-1}. \tag{21}$$

[1] Karup, *loc. cit.*, p. 83.

When $n=2$, we have six ordinates, y_{-2}, y_{-1}, y_0, y_1, y_2, y_3 and u_x is defined by the first five and v_x by the last five. Equations (6) and (11) make

$$(22) \qquad y_x = u_x + x^3(x-1)(a_0x + a_1)\Delta^5 y_{-2}/\underline{|4},$$

where $u_x = y_{-2} + (x+2)\Delta y_{-2} + \frac{(x+2)(x+1)}{\underline{|2}}\Delta^2 y_{-2}$

$$+ \frac{(x+2)(x+1)x}{\underline{|3}}\Delta^3 y_{-2} + \frac{(x+2)(x+1)x(x-1)}{\underline{|4}}\Delta^4 y_{-2}.$$

From equations (13) and (15) and (17)

$$\xi_x = a_0x^4 + a_1x^3, \qquad \mu_x = (x+1)\,x(x-2),$$

$$\xi_x' = 4\,a_0x^3 + 3\,a_1x^2, \quad \mu_x' = (x+1)x + x(x-2) + (x+1)(x-2).$$

$$\begin{aligned} a_0 + a_1 &= -2, \\ 4\,a_0 + 3\,a_1 &= -1. \end{aligned} \qquad \therefore a_0 = 5,\ a_1 = -7.$$

From (19) we have $c_x = \dfrac{x^2(5x-7)}{(x+1)(x-2)}$.

The above expression (22) for y_x readily transforms into the following formula arrived at by Karup:[1]

$$(23) \qquad y_x = y_0 + x\Delta y_0 + \frac{x(x-1)}{\underline{|2}}\Delta^2 y_{-1} + \frac{(x+1)x(x-1)}{\underline{|3}}\Delta^3 y_{-1}$$

$$+ \frac{(x+1)x(x-1)(x-2)}{\underline{|4}}\Delta^4 y_{-2} + \frac{x^3(x-1)(5x-7)}{\underline{|4}}\Delta^5 y_{-2}.$$

When $n=3$ we have eight ordinates, y_{-3}, y_{-2}, y_{-1}, y_0, y_1, y_2, y_3, y_4, and u_x is determined by the first seven and v_x by the last seven. Formulæ (6) and (11) make

$$(24) \qquad y_x = u_x + x^4(x-1)(a_0x^2 + a_1x + a_2)\Delta^7 y_{-3}/\underline{|6},$$

where

$$u_x = y_{-3} + (x+3)\Delta y_{-3} + \frac{(x+3)(x+2)}{\underline{|2}}\Delta^2 y_{-3}$$

$$+ \frac{(x+3)(x+2)(x+1)}{\underline{|3}}\Delta^3 y_{-3} + \frac{(x+3)(x+2)(x+1)x}{\underline{|4}}\Delta^4 y_{-3}$$

$$+ \frac{(x+3)(x+2)(x+1)x(x-1)}{\underline{|5}}\Delta^5 y_{-3}$$

$$+ \frac{(x+3)(x+2)(x+1)x(x-1)(x-2)}{\underline{|6}}\Delta^6 y_{-3}.$$

[1] Karup, *loc. cit.*, p. 83.

From relations (13), (15), and (17)

$$\xi_x = a_0x^6 + a_1x^5 + a_2x^4,$$
$$\xi_x' = 6\,a_0x^5 + 5\,a_1x^4 + 4\,a_2x^3,$$
$$\xi_x'' = 30\,a_0x^4 + 20\,a_1x^3 + 12\,a_2x^2,$$
$$\mu_x = (x+2)(x+1)x(x-2)(x-3)$$
$$= x^5 - 2x^4 - 7x^3 + 8x^2 + 12x,$$
$$\mu_x' = 5x^4 - 8x^3 - 21x^2 + 16x + 12,$$
$$\mu_x'' = 20x^3 - 24x^2 - 42x + 16,$$

whence

$$a_0 + a_1 + a_2 = 12,$$
$$6\,a_0 + 5\,a_1 + 4\,a_2 = 4,$$
$$30\,a_0 + 20\,a_1 + 12\,a_2 = -30,$$

and $$a_0 = 89,\ a_1 = -222,\ a_2 = 145.$$

The osculatory interpolation curve of the seventh degree thus becomes

$$(25) \qquad y_x = u_x + \frac{x^4(x-1)(89x^2 - 22x + 145)}{\lfloor 6}\Delta^7 y_{-3},$$

and by (19) it may be put in the form

$$y_x = (1 - c_x)u_x + c_x v_x,$$

where

$$c_x = \frac{x^3(89x^2 - 222x + 145)}{(x+2)(x+1)(x-2)(x-3)}.$$

The above expression for y_x may easily be transformed into the following:

$$(26) \quad y_x = y_0 + x\Delta y_0 + \frac{x(x-1)}{\lfloor 2}\Delta^2 y_{-1} + \frac{(x+1)x(x-1)}{\lfloor 3}\Delta^3 y_{-1}$$
$$+ \frac{(x+1)x(x-1)(x-2)}{\lfloor 4}\Delta^4 y_{-2} + \frac{(x+2)(x+1)x(x-1)(x-2)}{\lfloor 5}\Delta^5 y_{-2}$$
$$+ \frac{(x+2)(x+1)x(x-1)(x-2)(x-3)}{\lfloor 6}\Delta^6 y_{-3}$$
$$+ \frac{x^4(x-1)(89x^2 - 222x + 145)}{\lfloor 6}\Delta^7 y_{-3}.$$

An examination of the cases for $n = 1, 2, 3$, makes it easy to write down the general formula. A very interesting characteristic of the osculatory formula is that it differs from the corresponding ordinary interpolation formula only in the last term.

This suggests that by properly modifying or correcting the highest differences we might mechanically perform the osculatory interpolation; the details of this method were worked out by Karup in his paper[1] on mechanical graduation. Another point worth emphasizing is that the osculatory interpolation curve may be regarded as a blend or combination of the two curves u_x and v_x in the proportion of $1-c_x$ to c_x. The function c_x vanishes at $x=0$, and $1-c_x$ vanishes at $x=1$; it is continuous in the interval from 0 to 1 and may be regarded as the welding factor employed in this interval to weld together the two curves u_x and v_x to form the curve y_x. While in this discussion equal unit intervals have been assumed between the ordinates, it is not essential that the intervals be equal. And the theory admits of easy generalization in such manner that u_x and v_x may differ in degree, and osculatory interpolation curves of different degree may follow one another in succession. In short, many other conditions besides those peculiar to osculatory interpolation curves may be imposed and satisfied by properly choosing the constants $a_0, a_1, a_2, \cdots$ both as to number and value.[2]

In order to interpolate *four* values between $x=0$ and $x=1$, we form the leading differences in the usual manner and arrive at the following relations, to be employed with the osculatory formula where $n=2$.

$$
\begin{aligned}
\delta y_0 &= +.2\Delta y_{-2} + .32\Delta^2 y_{-2} + .088\Delta^3 y_{-2} - .0176\Delta^4 y_{-2} + .0016\Delta^5 y_{-2} \\
\delta^2 y_0 &= +.04\Delta^2 y_{-2} + .048\Delta^3 y_{-2} + .0016\Delta^4 y_{-2} + .0048\Delta^5 y_{-2} \\
\delta^3 y_0 &= +.008\Delta^3 y_{-2} + .0064\Delta^4 y_{-2} - .0048\Delta^5 y_{-2} \\
\delta^4 y_0 &= +.0016\Delta^4 y_{-2} - .0032\Delta^5 y_{-2} \\
\delta^5 y_0 &= +.0080\Delta^5 y_{-2}
\end{aligned}
\tag{27}
$$

APPLICATION TO CENSUS DATA.

We are now ready to apply the osculatory method to the data of Table I. The formula employed is (22). Since the population and deaths are given for quinquennial age groups they may be considered as being represented graphically by areas. For this reason the areas are summed from the highest age back to age x and a new table of the functions T_x and l_x is formed and these values become the subject of interpolation. T_x then denotes the population aged x and over, l_x the number of deaths within the calendar year of persons aged x and over.

[1] Karup, *loc. cit.*, p. 84. [2] Lidstone, *loc. cit.*, p. 284.

TABLE II.—POPULATION AND DEATHS OF PERSONS AGED x AND OVER.

AGGREGATE MALES, REGISTRATION STATES.

AGE.	POPULATION JUNE 1, 1900.	DEATHS, 1900.	AGE.	POPULATION JUNE 1, 1900.	DEATHS, 1900.
x	T_x	l_x	x	T_x	l_x
5	8,916,354	121,760	55	1,101,913	58,982
10	7,918,070	117,072	60	760,116	50,507
15	7,006,110	114,426	65	484,593	41,174
20	6,124,326	110,140	70	283,469	31,049
25	5,208,766	103,728	75	142,764	20,828
30	4,296,932	96,557	80	57,738	11,480
35	3,464,973	89,299	85	17,382	4,689
40	2,708,922	81,628	90	3,678	1,311
45	2,063,477	74,146	95	681	274
50	1,546,064	66,916	100	130	54

The effect of interpolating T_x and l_x by the osculatory method is to reproduce the values given in Table II and thus leave the totals in each quinquennial age group in Table I unchanged. It thus happens that the irregularities in the original data due to crowding at the ages which are multiples of five are eliminated and a smooth curve obtained without altering the quinquennial aggregates. This, of course, is precisely what is desired. There are various methods of graduating the data given in Table II, among which I have chosen that denoted by King[1] as Construction A. This consists in finding by interpolation of T_x and l_x the values of L_x and d_x for each age. The latter lead directly by calculation to the value of the death rate q_x.

To apply the fifth difference osculatory interpolation formula (22) to T_x the differences are formed as shown in Table III. These *quinquennial* differences are denoted by Δ, Δ^2, etc. The leading *unit* differences, denoted by δ, δ^2, etc., are then found by employing the multipliers at the foot of Table III in connection with the quinquennial differences in this same table in accordance with the relations (27). Table IV exhibits these leading unit differences.

[1] King, *loc. cit.*, p. 253.

TABLE III.—QUINQUENNIAL DIFFERENCES OF POPULATION AGED x AND OVER.

x	ΔT_x	$\Delta^2 T_x$	$\Delta^3 T_x$	$\Delta^4 T_x$	$\Delta^5 T_x$
5	−998,284	+ 86,324	−56,148	− 7,804	+109,258
10	−911,960	+ 30,176	−63,952	+101,454	− 62,807
15	−881,784	− 33,776	+37,502	+ 38,647	−118,763
20	−915,560	+ 3,726	+76,149	− 80,116	+118,781
25	−911,834	+ 79,875	− 3,967	+ 38,665	− 55,937
30	−831,959	+ 75,908	+34,698	− 17,272	− 54,924
35	−756,051	+110,606	+17,426	− 72,196	+156,058
40	−645,445	+128,032	−54,770	+ 83,862	−149,034
45	−517,413	+ 73,262	+29,092	− 65,172	+109,377
50	−444,151	+102,354	−36,080	+ 44,205	− 66,310
55	−341,797	+ 66,274	+ 8,125	− 22,105	+ 31,345
60	−275,523	+ 74,399	−13,980	+ 9,240	− 15,509
65	−201,124	+ 60,419	− 4,740	− 6,269	− 740
70	−140,705	+ 55,679	−11,009	− 7,009	+ 9,082
75	− 85,026	+ 44,670	−18,018	+ 2,073	+ 5,611
80	− 40,356	+ 26,652	−15,945	+ 7,684	
85	− 13,704	+ 10,707	− 8,261		
90	− 2,997	+ 2,446			
95	− 551				

Multipliers of above Δ's to form leading unit differences.

$\delta^5 T_{x+10}$					+.0080
$\delta^4 T_{x+10}$				+.0016	−.0032
$\delta^3 T_{x+10}$			+.0080	+.0064	−.0048
$\delta^2 T_{x+10}$		+.0400	+.0480	+.0016	+.0048
$\delta\ T_{x+10}$	+.2000	+.3200	+.0880	−.0176	+.0016

TABLE IV.—LEADING UNIT DIFFERENCES OF POPULATION AGED x AND OVER.

x	$\delta^5 T_x$	$\delta^4 T_x$	$\delta^3 T_x$	$\delta^2 T_x$	δT_x
15	+ 874.0640	−362.1120	−1023.5680	+1269.8080	−176661.9808
20	− 502.4560	+363.3088	+ 439.1632	−2001.8032	−180249.5376
25	− 950.1040	+441.8768	+1117.4192	− 59.1712	−184735.1520
30	+ 950.2480	−508.2848	− 473.6992	+4246.1552	−173618.4768
35	− 447.4960	+240.8624	+ 484.2176	+2797.9504	−157925.8992
40	− 439.3920	+148.1216	+ 430.6784	+4410.5536	−138831.7072
45	+1248.4640	−614.8992	−1071.7248	+5894.2528	−112762.4496
50	−1192.2720	+611.0880	+ 813.9200	+1911.1360	− 94652.9456
55	+ 875.0160	−454.2816	− 709.3744	+4747.6304	− 76156.6336
60	− 530.4800	+282.9200	+ 312.5600	+2114.7600	− 60136.0640
65	+ 250.7600	−135.6720	− 226.9280	+3156.0480	− 45997.5200
70	− 124.0720	+ 64.4128	+ 21.7392	+2245.2608	− 32714.5984
75	− 5.9200	− 7.6624	− 74.4896	+2175.6576	− 21198.6896
80	+ 72.6560	− 40.2768	− 176.5232	+1731.1072	− 11154.6224
85	+ 44.8880	− 14.6384	− 157.8096	+ 952.1856	− 4323.8912

By employing these leading unit differences we are now able to derive the T_x table for each age from 15 to 90. It is not necessary, however, to go beyond the δT_x column in this computation, as we desire only the population or L_x table for each age. The numerical check on the work appears in the summation of the five first differences in each interpolation group, for this sum should be equal to the difference between the corresponding pair of quinquennial values of T_x taken from Table III. This process is illustrated in the following calculation for the two age groups 15–19 and 35–39.

FIRST NUMERICAL ILLUSTRATION.

Age Group 15–19.		Age Group 35–39.	
	+874.0640		−447.4960
	−362.1120		+240.8624
	−1023.5680		+484.2176
	+1269.8080		+2797.9504
	−176661.9808		−157925.8992
	+511.9520		−206.6336
	−1385.6800		+725.0800
	+246.2400		+3282.1680
	−175392.1728		−155127.9488
	−873.7280		+518.4464
	−1139.4400		+4007.2480
	−175145.9328		−151845.7808
	−2013.1680		+ 4525.6944
	−176285.3728		−147838.5328
	−178298.5408		−143312.8384
15	176661.9808	35	157925.8992
16	175392.1728	36	155127.9488
17	175145.9328	37	151845.7808
18	176285.3728	38	147838.5328
19	178298.5408	39	143312.8384
15–19	881784.0000 $= T_{15} - T_{20}$	35–39	756051.0000 $= T_{35} - T_{40}$

Starting with the five leading unit differences opposite age 15 in Table IV, we form by successive addition in the usual manner, as appearing in the first column of the numerical illustration, the values of δT_x for the four ages between 15 and 20; a check on the correctness of the numerical work in this column and the calculation of the five leading unit differences from the five leading quinquennial differences *of an age ten years younger* is the fact that the sum, $-(\delta T_{15} + \delta T_{16} + \delta T_{17} + \delta T_{18} + \delta T_{19}) = 881{,}784$, agrees with the negative quinquennial difference, $-\Delta T_{15} = T_{15} - T_{20} = 881{,}784$, taken from Table III. This check, however, does not insure the correctness of the leading quinquennial differences employed in the calculation, for $\Delta^4 T_{x-10}$ and $\Delta^5 T_{x-10}$ in particular may be chosen at pleasure and the above check satisfied if the numerical work is correct. This is made apparent by an examination of the following numerical illustration, which exhibits the process of deriving the interpolated values by the successive addition of the leading unit differences, which in turn are expressed in terms of the leading quinquennial differences of an age ten years younger. The coefficients of the respective quinquennial differences appear in the columns below them.

SECOND NUMERICAL ILLUSTRATION.

	ΔT_{x-10}	$\Delta^2 T_{x-10}$	$\Delta^3 T_{x-10}$	$\Delta^4 T_{x-10}$	$\Delta^5 T_{x-10}$
$\delta^5 T_x$	.0000	.0000	.0000	.0000	+.0080
$\delta^4 T_x$	.0000	.0000	.0000	+.0016	−.0032
$\delta^3 T_x$	.0000	.0000	+.0080	+.0064	−.0048
$\delta^2 T_x$	.0000	+.0400	+.0480	+.0016	+.0048
δT_x	+.2000	+.3200	+.0880	−.0176	+.0016
	.0000	.0000	.0000	+.0016	+.0048
	.0000	.0000	+.0080	+.0080	−.0080
	.0000	+.0400	+.0560	+.0080	.0000
	+.2000	+.3600	+.1360	−.0160	+.0064
	.0000	.0000	+.0080	+.0096	−.0032
	.0000	+.0400	+.0640	+.0160	−.0080
	+.2000	+.4000	+.1920	−.0080	+.0064
	.0000	+.0400	+.0720	+.0256	−.0112
	+.2000	+.4400	+.2560	+.0080	−.0016
	+.2000	+.4800	+.3280	+.0336	−.0128
δT_x	+.2000	+.3200	+.0880	−.0176	+.0016
δT_{x+1}	+.2000	+.4800	+.3280	+.0336	−.0128
δT_{x+2}	+.2000	+.4400	+.2560	+.0080	−.0016
δT_{x+3}	+.2000	+.4000	+.1920	−.0080	+.0064
δT_{x+4}	+.2000	+.3600	+.1360	−.0160	+.0064
$\sum_x^{x+4} \delta T_x$	+1.0000	+2.0000	+1.0000	.0000	.0000

The final summation in the above numerical illustration, which corresponds to our so-called check in the first numerical illustration, appears to be merely a statement of the known relation

$$\sum_x^{x+4} \delta T_x = \Delta T_{x-10} + 2\,\Delta^2 T_{x-10} + \Delta^3 T_{x-10} = \Delta T_x,$$

which, indeed, we might have predicted in advance. Since the multipliers or coefficients of $\Delta^4 T_{x-10}$ and $\Delta^5 T_{x-10}$ combine in such manner in this process as to vanish, it is evident that these two differences may be taken arbitrarily, and the relation

$$\sum_x^{x+4} \delta T_x = \Delta T_x = T_{x+5} - T_x$$

will still be satisfied if the numerical work is correct.

The general relation would be as follows:

$$\Delta T_x = \Delta T_{x-5n} + n\Delta^2 T_{x-5n} + \frac{n(n-1)}{\underline{|2}} \Delta^3 T_{x-5n} + \cdots + \Delta^{n+1} T_{x-5n},$$

from which it appears that the last n quinquennial differences,

$$\Delta^{n+2}T_{x-5n},\ \Delta^{n+3}T_{x-5n},\ \cdots\ \Delta^{2n+1}T_{x-5n}$$

may be arbitrarily chosen in the general case, $n = n$, without disturbing the relation

$$\sum_{x}^{x+4} \delta T_x = \Delta T_x.$$

It is therefore evident that the differences in Table III must be first tested by some other formula, such as

$$\sum_{1}^{n} \Delta u_n = u_{n+1} - u_1,$$

and their correctness absolutely assured before we employ them to form the leading unit differences in Table IV and proceed to the calculation of the interpolated values of δT_x. It would be a useless labor and a fallacious test to add the respective columns of leading quinquennial differences in Table III and apply to these sums the multipliers at the foot of the table and check the same against the sums of the corresponding columns of the leading unit differences in Table IV. I have dwelt at some length on this point because King[1] offers it as "a complete final check on the whole work," and with the hope that this word of caution may be of assistance to others undertaking calculations of this nature.[2]

Returning to the examination of the first numerical illustration, we observe that the aggregate population in the quinquennial age group 15–19 remains undisturbed and the redistribution is made in such manner that the figures form the first differences, apart from the negative sign, of a function T_x which has been made the subject of an osculatory interpolation by fifth differences. The following table exhibits the enumerated and graduated population for each age from fifteen to eighty-nine, both inclusive.

[1] King, *loc. cit.*, p. 244.

[2] After this discussion was written I found that Sprague had noted the same thing in connection with the fifth difference equation, where $n = 3$. Sprague, *loc. cit.*, p. 284.

TABLE V.—POPULATION, ENUMERATED AND GRADUATED.

AGGREGATE MALES, REGISTRATION STATES.

AGE.	CENSUS, JUNE 1, 1900.	GRADUATED.	AGE.	CENSUS, JUNE 1, 1900.	GRADUATED.
15	177,666	176,662	45	131,636	112,762
16	180,668	175,392	46	96,579	106,868
17	174,772	175,146	47	94,312	102,046
18	176,597	176,285	48	100,247	98,910
19	172,081	178,299	49	94,639	96,827
15–19	881,784	881,784	45–49	517,413	517,413
20	178,264	180,250	50	126,236	94,653
21	180,837	182,251	51	75,903	92,742
22	182,552	183,814	52	87,851	90,017
23	183,814	184,574	53	76,617	85,866
24	190,093	184,671	54	77,544	80,873
20–24	915,560	915,560	50–54	444,151	444,151
25	193,570	184,735	55	85,830	76,157
26	180,066	184,794	56	72,184	71,409
27	181,020	183,736	57	62,508	67,371
28	189,633	181,119	58	62,175	64,496
29	167,545	177,450	59	59,100	62,364
25–29	911,834	911,834	55–59	341,797	341,797
30	215,807	173,619	60	79,876	60,136
31	141,323	169,372	61	45,214	58,021
32	167,874	165,600	62	52,610	55,594
33	154,650	162,809	63	50,989	52,571
34	152,305	160,559	64	46,834	49,201
30–34	831,959	831,959	60–64	275,523	275,523
35	176,535	157,926	65	52,700	45,998
36	143,012	155,128	66	40,688	42,841
37	135,897	151,846	67	39,672	39,912
38	157,661	147,838	68	35,659	37,346
39	142,946	143,313	69	32,405	35,027
35–39	756,051	756,051	65–69	201,124	201,124
40	181,577	138,832	70	40,103	32,715
41	108,908	134,421	71	25,519	30,469
42	132,531	129,580	72	28,131	28,202
43	112,420	124,160	73	24,757	25,849
44	110,009	118,452	74	22,195	23,470
40–44	645,445	645,445	70–74	140,705	140,705

TABLE V.— POPULATION, ENUMERATED AND GRADUATED.— *Continued.*

AGGREGATE MALES, REGISTRATION STATES.

AGE.	CENSUS, JUNE 1, 1900.	GRADUATED.	AGE.	CENSUS, JUNE 1, 1900.	GRADUATED.
75	23,496	21,199	85	4,420	4,324
76	18,599	19,023	86	3,068	3,372
77	15,852	16,922	87	2,675	2,577
78	14,640	14,903	88	2,052	1,955
79	12,439	12,979	89	1,489	1,476
75–79	85,026	85,026	85–89	13,704	13,704
80	12,680	11,155			
81	8,750	9,423			
82	7,514	7,869			
83	6,030	6,531			
84	5,382	5,378			
80–84	40,356	40,356			

It will be noticed that there is a strong tendency to mass about ages which are multiples of five, especially in the section of the table between ages 30 and 60. About age 30, for example, the adjustment amounts to a considerable change in the enumerated data, the corrections appearing as follows: —

AGE.	CENSUS.	GRADUATED.	CORRECTION.
28	189,633	181,119	− 8,514
29	167,545	177,450	+ 9,905
30	215,807	173,619	−42,188
31	141,323	169,372	+28,049
32	167,874	165,600	− 2,274

While we are impressed with the remarkable smoothness of the results obtained by the osculatory interpolation, we are compelled to regard these adjustments involving such extensive shifting of the figures as open to the objection of being artificial, perhaps in a high degree. Although we are certain that the enumerated population at age 30 is too great, it would be going too far to assert that this method of adjustment is the one most likely to give the approximately true figures.

We have next to deal with the vital statistics. The data given in the l_x column in Table II is now treated in the same manner as T_x in the preceding paragraphs. In fact, to T_x and L_x in that discussion, correspond l_x and d_x, respectively, in what follows.

46710

Table VI gives the quinquennial differences of l_x, together with the multipliers to be used in forming the leading unit differences to be employed to obtain the interpolated values of d_x for each age. Table VII exhibits these calculated leading unit differences.

TABLE VI.—QUINQUENNIAL DIFFERENCES OF DEATHS AT AGE x AND OVER.

x	Δl_x	$\Delta^2 l_x$	$\Delta^3 l_x$	$\Delta^4 l_x$	$\Delta^5 l_x$
5	− 4,688	+2,042	−3,682	+3,196	−1,343
10	− 2,646	−1,640	− 486	+1,853	−2,548
15	− 4,286	−2,126	+1,367	− 695	− 303
20	− 6,412	− 759	+ 672	− 998	+1,926
25	− 7,171	− 87	− 326	+ 928	−1,467
30	− 7,258	− 413	+ 602	− 539	− 480
35	− 7,671	+ 189	+ 63	−1,019	+2,138
40	− 7,482	+ 252	− 956	+1,119	−1,599
45	− 7,230	− 704	+ 163	− 480	+ 863
50	− 7,934	− 541	− 317	+ 383	+ 247
55	− 8,475	− 858	+ 66	+ 630	− 357
60	− 9,333	− 792	+ 696	+ 273	+ 442
65	−10,125	− 96	+ 969	+ 715	−1,543
70	−10,221	+ 873	+1,684	− 828	−1,100
75	− 9,348	+2,557	+ 856	−1,928	+1,476
80	− 6,791	+3,413	−1,072	− 452	
85	− 3,378	+2,341	−1,524		
90	− 1,037	+ 817			
95	− 220				

Multipliers of above Δ's to form leading unit differences.

$\delta^5 l_{x+10}$					+.0080
$\delta^4 l_{x+10}$				+.0016	−.0032
$\delta^3 l_{x+10}$			+.0080	+.0064	−.0048
$\delta^2 l_{x+10}$		+.0400	+.0480	+.0016	+.0048
$\delta\, l_{x+10}$	+.2000	+.3200	+.0880	−.0176	+.0016

TABLE VII.—LEADING UNIT DIFFERENCES OF DEATHS AGED x AND OVER.

x	$\delta^5 l_x$	$\delta^4 l_x$	$\delta^3 l_x$	$\delta^2 l_x$	δl_x
15	−10.7440	+ 9.4112	− 2.5552	− 96.3888	− 666.5744
20	−20.3840	+11.1184	+20.2016	− 98.1936	−1133.4576
25	− 2.4240	− 0.1424	+ 7.9424	− 21.9904	−1405.4768
30	+15.4080	− 7.7600	−10.2560	+ 9.5440	−1445.4976
35	−11.7360	+ 6.1792	+10.3728	− 24.6848	−1509.4080
40	− 3.8400	+ 0.6736	+ 3.6704	+ 9.2096	−1522.0656
45	+17.1040	− 8.4720	−16.2800	+ 19.2160	−1446.8208
50	−12.7920	+ 6.9072	+ 7.1888	− 41.6928	−1522.1408
55	+ 6.9040	− 3.5296	− 5.9104	− 16.9616	−1647.1072
60	+ 1.9760	− 0.1776	− 1.2704	− 35.0576	−1794.1616
65	− 2.8560	+ 2.1504	+ 6.2736	− 31.8576	−1975.4112
70	+ 3.5360	− 0.9776	+ 5.1936	+ 4.2864	−2062.8896
75	−12.3440	+ 6.0816	+19.7344	+ 36.4096	−1985.5008
80	− 8.8000	+ 2.1952	+13.4528	+109.1472	−1603.8352
85	+11.8080	− 7.8080	−12.5760	+147.3680	− 939.7376

The third numerical illustration, which appears below, exhibits the process of calculating the values of d_x for the age groups 15–19 and 35–39. A check on the numerical work in the first column and the calculation of the five leading unit differences opposite age 15 in Table VII from the five leading quinquennial differences opposite age 5 in Table VI is obtained because $-(\delta l_{15} + \delta l_{16} + \delta l_{17} + \delta l_{18} + \delta l_{19}) = 4286$ agrees with the negative quinquennial difference, $-\Delta l_{15} = l_{15} - l_{20} = 4286$, taken from Table VI. This check does not insure the correctness of the leading quinquennial differences opposite age 5 in Table VI.

THIRD NUMERICAL ILLUSTRATION.

	Age Group 15–19.		Age Group 35–39.
	− 10.7440		− 11.7360
	+ 9.4112		+ 6.1792
	− 2.5552		+ 10.3728
	− 96.3888		− 24.6848
	−666.5744		−1509.4080
	− 1.3328		− 5.5568
	+ 6.8560		+ 16.5520
	− 98.9440		− 14.3120
	−762.9632		−1534.0928
	+ 5.5232		+ 10.9952
	− 92.0880		+ 2.2400
	−861.9072		−1548.4048
	− 86.5648		+ 13.2352
	−953.9952		−1546.1648
	−1040.5600		−1532.9296
15	666.5744	35	1509.4080
16	762.9632	36	1534.0928
17	861.9072	37	1548.4048
18	953.9952	38	1546.1648
19	1040.5600	39	1532.9296
15–19	$4286.0000 = l_{15} - l_{20}$	35–39	$7671.0000 = l_{35} - l_{40}$

The next table furnishes a comparison between the observed and graduated data with respect to deaths. It was made possible through the courtesy of Dr. Cressy L. Wilbur, Chief Statistician of the Division of Vital Statistics, who kindly furnished me with some of the advance sheets of the volume on Mortality Statistics for 1908. The reported deaths for the registration states in the calendar year 1900 are here given for each age, thus affording a comparison of deaths similar to that set forth on population in Table V. The process of osculatory interpolation having been applied to the quinquennial values of the function l_x, the deaths at each age, apart from the algebraic sign, appear as the first differences of this function, consequently the aggregates in the quinquennial age groups are unaltered in value but subjected to a new distribution of the individual ages within the age group. The general effect, as seen by an examination of Table VIII, is to smooth out the irregularities due to massing at the ages which are multiples of five. This table, together with Tables IX, X, and XI, appears at the end of the paper.

Table IX brings together the graduated population and deaths for each age, and the remaining column exhibits the corresponding death rates. The following formula was employed in computing the death rate: —

$$q_x = \frac{2\,d_x}{2\,L_x + d_x} = \frac{d_x}{L_x + \frac{1}{2}\,d_x}.$$

It is based upon the assumption that deaths occur uniformly throughout the year, so that at the beginning of the year there are $L_x + \frac{1}{2}\,d_x$ persons aged x, among whom d_x deaths occur at uniform intervals during the year, leaving a population of L_x in the middle and $L_x - \frac{1}{2}\,d_x$ at the end of the year. The death rate is given to six places of decimals.

Table X exhibits the death rate cut down to five places of decimals, and the third difference of the death rate multiplied by 100,000. This table is designed to indicate the extreme smoothness of the graduation of the death rate. It is possible that by further refinements some little improvement might be made in this particular, but I am inclined to think it would be unimportant and that the rates as they now stand closely represent the average actual conditions in the registration area of the United States.

Table XI is added to facilitate a comparison between the census death rates in this country and England. The New York City table, which was graduated by Roche[1] from census data, is exhibited to show the striking excess in the death rate in New York City over the average of the registration states. It is not my purpose, however, to discuss any of these features in this paper. The English Life Table, No. 6, was graduated by King.[2] A comparison shows that from ages 15 to 37 the death rate in this country is higher than in England, but after age 37 the reverse is true.

The death rates for ages preceding 15 and following 89 have not been considered in this paper, but I hope at no distant time to fill in this gap and construct the complete mortality table.

[1] John F. Roche, An Investigation into the Mortality Rates of the City of New York, Transactions Actuarial Society of America, Vol. 7, p. 435.

[2] King, *loc. cit.*, p. 274.

TABLE VIII.—COMPARISON OF REPORTED AND GRADUATED DEATHS.

AGGREGATE MALES, REGISTRATION STATES.

AGE.	REPORTED 1900.	GRADUATED.	AGE.	REPORTED 1900.	GRADUATED.	AGE.	REPORTED 1900.	GRADUATED.
15	603	667	40	1,976	1,522	65	2,292	1,975
16	754	763	41	1,194	1,513	66	1,830	2,007
17	847	862	42	1,588	1,500	67	2,004	2,033
18	976	954	43	1,394	1,483	68	2,063	2,050
19	1,106	1,040	44	1,330	1,464	69	1,936	2,060
15–19	4,286	4,286	40–44	7,482	7,482	65–69	10,125	10,125
20	1,108	1,133	45	1,748	1,447	70	2,436	2,063
21	1,249	1,232	46	1,284	1,428	71	1,728	2,059
22	1,290	1,310	47	1,256	1,425	72	2,066	2,049
23	1,306	1,356	48	1,487	1,446	73	1,987	2,035
24	1,459	1,381	49	1,455	1,484	74	2,004	2,015
20–24	6,412	6,412	45–49	7,230	7,230	70–74	10,221	10,221
25	1,415	1,405	50	1,985	1,522	75	2,191	1,986
26	1,398	1,427	51	1,258	1,564	76	1,983	1,949
27	1,439	1,442	52	1,605	1,598	77	1,721	1,893
28	1,518	1,448	53	1,496	1,619	78	1,859	1,811
29	1,401	1,449	54	1,590	1,631	79	1,594	1,709
25–29	7,171	7,171	50–54	7,934	7,934	75–79	9,348	9,348
30	1,702	1,445	55	1,903	1,647	80	1,813	1,604
31	1,174	1,436	56	1,754	1,664	81	1,341	1,495
32	1,584	1,437	57	1,626	1,687	82	1,356	1,372
33	1,449	1,455	58	1,685	1,719	83	1,146	1,234
34	1,349	1,485	59	1,507	1,758	84	1,135	1,086
30–34	7,258	7,258	55–59	8,475	8,475	80–84	6,791	6,791
35	1,755	1,509	60	2,237	1,794	85	951	940
36	1,410	1,534	61	1,503	1,829	86	766	792
37	1,362	1,549	62	1,791	1,866	87	663	658
38	1,732	1,546	63	1,868	1,903	88	550	543
39	1,412	1,533	64	1,934	1,941	89	448	445
35–39	7,671	7,671	60–64	9,333	9,333	85–89	3,378	3,378

TABLE IX.—ADJUSTED POPULATION, DEATHS, AND DEATH RATES.

AGGREGATE MALES, REGISTRATION STATES.

AGE.	POPULATION JUNE 1, 1900.	DEATHS, 1900.	DEATH RATE.	AGE.	POPULATION JUNE 1, 1900.	DEATHS, 1900.	DEATH RATE.
x	L_x	d_x	q_x	x	L_x	d_x	q_x
15	176,662	667	.003768	55	76,157	1,647	.021395
16	175,392	763	.004341	56	71,409	1,664	.023034
17	175,146	862	.004910	57	67,371	1,687	.024731
18	176,285	954	.005397	58	64,496	1,719	.026302
19	178,299	1,040	.005816	59	62,364	1,758	.027798
20	180,250	1,133	.006266	60	60,136	1,794	.029394
21	182,251	1,232	.006737	61	58,021	1,829	.031034
22	183,814	1,310	.007101	62	55,594	1,866	.033011
23	184,574	1,356	.007320	63	52,571	1,903	.035555
24	184,671	1,381	.007450	64	49,201	1,941	.038687
25	184,735	1,405	.007577	65	45,998	1,975	.042034
26	184,794	1,427	.007692	66	42,841	2,007	.045775
27	183,736	1,442	.007818	67	39,912	2,033	.049672
28	181,119	1,448	.007963	68	37,346	2,050	.053426
29	177,450	1,449	.008132	69	35,027	2,060	.057132
30	173,619	1,445	.008288	70	32,715	2,063	.061132
31	169,372	1,436	.008443	71	30,469	2,059	.065368
32	165,600	1,437	.008640	72	28,202	2,049	.070108
33	162,809	1,455	.008897	73	25,849	2,035	.075745
34	160,559	1,485	.009206	74	23,470	2,015	.082320
35	157,926	1,509	.009510	75	21,199	1,986	.089492
36	155,128	1,534	.009840	76	19,023	1,949	.097462
37	151,846	1,549	.010149	77	16,922	1,893	.105941
38	147,838	1,546	.010403	78	14,903	1,811	.114559
39	143,313	1,533	.010640	79	12,979	1,709	.123541
40	138,832	1,522	.010903	80	11,155	1,604	.134147
41	134,421	1,513	.011193	81	9,423	1,495	.146994
42	129,580	1,500	.011509	82	7,869	1,372	.160374
43	124,160	1,483	.011873	83	6,531	1,234	.172636
44	118,452	1,464	.012284	84	5,378	1,086	.183415
45	112,762	1,447	.012751	85	4,324	940	.196078
46	106,868	1,428	.013274	86	3,372	792	.210191
47	102,046	1,425	.013867	87	2,577	658	.226428
48	98,910	1,446	.014513	88	1,955	543	.243881
49	96,827	1,484	.015210	89	1,476	445	.261996
50	94,653	1,522	.015952				
51	92,742	1,564	.016723				
52	90,017	1,598	.017596				
53	85,866	1,619	.018679				
54	80,873	1,631	.019966				

TABLE X.—AMERICAN LIFE TABLE. AGGREGATE MALES.

REGISTRATION STATES. UNITED STATES CENSUS, 1900.

AGE.	q_x	$10^5\Delta^3 q_x$	AGE.	q_x	$10^5\Delta^3 q_x$
15	.00377	− 8	55	.02140	− 20
16	.00434	+ 1	56	.02303	+ 10
17	.00491	+10	57	.02473	+ 12
18	.00540	− 1	58	.02630	− 4
19	.00582	−13	59	.02780	+ 29
20	.00627	− 3	60	.02939	+ 23
21	.00674	+ 5	61	.03103	+ 1
22	.00710	+ 9	62	.03301	− 37
23	.00732	− 2	63	.03556	+ 20
24	.00745	+ 4	64	.03869	− 27
25	.00758	− 1	65	.04203	− 27
26	.00769	+ 2	66	.04578	+ 7
27	.00782	− 4	67	.04967	+ 36
28	.00796	0	68	.05343	− 6
29	.00813	+ 6	69	.05713	+ 26
30	.00829	+ 1	70	.06113	+ 40
31	.00844	− 1	71	.06537	+ 3
32	.00864	− 6	72	.07011	− 33
33	.00890	+ 4	73	.07575	+ 20
34	.00921	− 5	74	.08232	− 29
35	.00951	− 4	75	.08949	− 37
36	.00984	+ 5	76	.09746	+ 22
37	.01015	+ 3	77	.10594	+127
38	.01040	+ 1	78	.11456	+ 60
39	.01064	0	79	.12354	−169
40	.01090	+ 1	80	.13415	−165
41	.01119	+ 1	81	.14699	− 38
42	.01151	+ 1	82	.16037	+337
43	.01187	− 1	83	.17264	− 43
44	.01228	+ 3	84	.18342	+ 68
45	.01275	− 4	85	.19608	− 92
46	.01327	+ 2	86	.21019	− 54
47	.01387	− 2	87	.22643	
48	.01451	− 1	88	.24388	
49	.01521	+ 8	89	.26200	
50	.01595	+ 9			
51	.01672	+ 1			
52	.01760	− 7			
53	.01868	+ 6			
54	.01997	−19			

TABLE XI.—COMPARATIVE TABLE OF DEATH RATES, MALES.

Age.	Registration States.	New York City.	English Life Table No. 6.	Age.	Registration States.	New York City.	English Life Table No. 6.
x	q_x	q_x	q_x	x	q_x	q_x	q_x
15	.00377	.00392	.00303	55	.02140	.03284	.02558
16	.00434	.00430	.00322	56	.02303	.03477	.02726
17	.00491	.00465	.00351	57	.02473	.03676	.02917
18	.00540	.00510	.00387	58	.02630	.03926	.03127
19	.00582	.00581	.00425	59	.02780	.04203	.03351
20	.00627	.00664	.00459	60	.02939	.04421	.03583
21	.00674	.00751	.00487	61	.03103	.04714	.03823
22	.00710	.00814	.00508	62	.03301	.05034	.04069
23	.00732	.00864	.00522	63	.03556	.05460	.04328
24	.00745	.00908	.00537	64	.03869	.05926	.04617
25	.00758	.00966	.00555	65	.04203	.06328	.04949
26	.00769	.01019	.00575	66	.04578	.06584	.05328
27	.00782	.01058	.00598	67	.04967	.07075	.05765
28	.00796	.01102	.00624	68	.05343	.07569	.06251
29	.00813	.01152	.00653	69	.05713	.08068	.06763
30	.00829	.01203	.00685	70	.06113	.08547	.07279
31	.00844	.01240	.00720	71	.06537	.09067	.07801
32	.00864	.01293	.00761	72	.07011	.09488	.08319
33	.00890	.01332	.00805	73	.07575	.09939	.08854
34	.00921	.01387	.00853	74	.08232	.10558	.09450
35	.00951	.01442	.00903	75	.08949	.11236	.10145
36	.00984	.01491	.00955	76	.09746	.11765	.10948
37	.01015	.01545	.01009	77	.10594	.12639	.11878
38	.01040	.01631	.01065	78	.11456	.14159	.12943
39	.01064	.01699	.01122	79	.12354	.15385	.14136
40	.01090	.01736	.01179	80	.13415	.16561	.15457
41	.01119	.01787	.01236	81	.14699	.18040	.16923
42	.01151	.01843	.01291	82	.16037	.19209	.18552
43	.01187	.01925	.01347	83	.17264	.20225	.20352
44	.01228	.02065	.01408	84	.18342	.21772	.22329
45	.01275	.02155	.01476	85	.19608	.22930	.24491
46	.01327	.02272	.01552	86	.21019	.24123	.26844
47	.01387	.02382	.01637	87	.22643	.25372	.29394
48	.01451	.02493	.01731	88	.24388	.26659	.32143
49	.01521	.02604	.01831	89	.26200	.28000	.35091
50	.01595	.02693	.01935				
51	.01672	.02788	.02043				
52	.01760	.02903	.02156				
53	.01868	.03034	.02276				
54	.01997	.03179	.02408				

THE AGE RETURNS OF THE TWELFTH CENSUS (1900).

By WILLIAM B. BAILEY, *Professor of Economics, Yale University,* and JULIUS H. PARMELEE, *Special Agent, Bureau of the Census.*

Beginning with 1850, inquiry has been made at each census regarding the exact age in years of every person in the United States. The inquiry has usually referred to age at last birthday, although in 1890 the census schedule asked for "age at nearest birthday." In the Twelfth Census (1900) the inquiry was divided into two parts, the first asking for month and year of birth, the second for age at last birthday. The purpose of the addition of the question regarding date of birth was to increase the degree of accuracy of the statistics of age.

The suggestion that date of birth be ascertained appeared originally in the population report of the Eleventh Census (1890). The instructions for that census had especially cautioned enumerators against accepting round numbers in answer to the age query. "Particular attention should be paid to this," the instructions had read,* "otherwise it will be found when the results are aggregated in this office that a much more than normal number of persons have been reported as 20, 25, 30, 35, 40, etc., years of age, and a much less than normal at 19, 21, 24, 26, 29, 31, etc." As a result of this caution, concentration on multiples of five was somewhat lessened in 1890, but as it was not wholly done away with, the suggestion was made,† that to return ages by exact day, month, and year of birth would still further reduce such concentration.

The inquiry regarding the month and year of birth which was introduced into the schedule for the Twelfth Census conformed to this suggestion; and when analyses of the age re-

*Wright and Hunt, History and Growth of the United States Census, p. 188.
†Report on Population, Census of 1890, Part II, p. XV.

turns of that census indicated a gain in accuracy and a decrease in concentration on multiples of 5,* both results were at once ascribed to the inclusion of this additional inquiry. Dr. Allyn A. Young, who made a special study of age statistics for the Twelfth Census, said: "The addition of the inquiry as to date of birth has increased the general accuracy of the age returns."† He believed also, that the decrease in the amount of concentration on multiples of 5 was due to the addition of this same inquiry, and even that "as high a degree of accuracy would have been obtained if the inquiry as to 'age at last birthday' had been omitted."‡

In considering the advisability of retaining both inquiries regarding age in the population schedule of the Thirteenth Census, the question arose whether the inclusion in 1900 of the inquiry respecting date of birth did or did not result in improving the age statistics of that year. To assist in reaching a decision on this point, a critical study of the question was initiated by the Census Bureau, leading to the results and conclusions presented in this paper.

A careful inspection of ten bound volumes of the original census schedules of 1900 was made, to learn how large was the number of cases in which the reported year of age was inconsistent with the reported date of birth. It was found that out of 130,000 enumerations taken by census agents in five rural counties (one each in Kansas, Mississippi, New Hampshire, Oregon, and Pennsylvania) and five large cities (Boston, Chicago, New Orleans, New York and San Francisco), there were 12,526 cases of disagreement between date of birth and age recorded, or nearly one in ten.

In a certain number of these cases of disagreement the other data recorded on the schedules furnished conclusive evidence, or at least created a very strong presumption, as to which of the two age returns was the more reliable, one return being

*The estimated coefficient of error in the age returns fell from 7.5 per cent. in 1890 to 3.4 per cent. in 1900, or more than one half. The excess per cent. of concentration on multiples of 5 was reduced from 31.3 per cent. in 1890 to 19.8 per cent. in 1900. See Supplementary Analysis, Census of 1900, pp. 135, 137.

†Supplementary Analysis, Census of 1900, p. 141.

‡Supplementary Analysis of 1900, p. 137. See also Report on Population, Census of 1900, Part II, p. XXXV.

consistent with the other record data and the other inconsistent. Typical cases of this sort are the following, which are taken from Volume 145 of the bound schedules of 1900 for New York State and Volume 19 for Illinois:

Case 1

A—Son, born December, 1889, age 11 years.

B—Daughter, born May, 1890, age 10 years.

The ages reported by the enumerator were probably correct, and the year of birth of A incorrect. Otherwise A would be but five months older than his sister B.

Case 2

C—Born August, 1861, age 8 years.

As the schedule reported C to be attending school, the age reported by the enumerator was doubtless correct, and the year of birth incorrect.

Case 3

D—Born 1849, age 60.

E—Son of D, born 1859, age 40.

Here again the reported age of D was in all probability correct, and not the year of his birth; otherwise he would be but 10 years older than his son E.

Case 4

F—Wife, born 1872, age 38.

F in 1900 had been married 20 years. Her age, therefore, was correctly reported as 38, and the year of birth assigned to her was unquestionably wrong.

Case 5

G—Daughter, born 1882, age 13.

The reported age and year of birth of G are in disagreement. As G's parents, aged 35 and 33, respectively, had been married but 14 years, it is almost beyond question that her age is reported accurately, and the year of her birth inaccurately.

Case 6

H—Daughter, born October, 1899, age 10.

H was reported as attending school, which would hardly have been possible had the reported date of her birth been correct.

Case 7

J—Husband, born 1877, age 33.

J had been married 13 years and was the father of a boy of 11. His reported age is clearly more reliable than the reported year of his birth.

Case 8

K—Wife, born 1852, age 27.

Here the year of birth seems more trustworthy than the reported age in years. K was a foreign-born woman who had been in the country 28 years, and had been married 22 years of that time. Her reported age therefore, is evidently incorrect.

Of the cases selected as examples and cited above seven indicate an incorrect return of the year of birth and one an incorrect return of the year of age. In a count of 50 similar cases taken at random, or rather in the order in which they appeared in the bound volumes of the schedules consulted for this purpose, it was found that all but 6 indicated very conclusively that the return of age was correct, rather than that of year of birth. The fact that this proportion obtained in those cases where one could check the comparative accuracy of the two returns created a very strong presumption that in other cases, where there was no such check, the return of age was much more likely to be accurate than that of date of birth.

Upon a further inspection of the 12,526 cases of disagreement which had been checked off in the schedules, it was found that in more than two-thirds of these cases, the difference between the year of age return and the date of birth return was exactly one year, and that in every such case the person, according to the reported date of birth, was just one year

younger than his reported year of age. This typical relationship between the two returns in cases of disagreement could not have been accidental. It was evident there must be some causal connection — some explanation of the tendency to make the date of birth return one year younger than the year of age return. Further light was thrown upon the question when it was found that in all these cases almost without exception the birth month was one of the last seven months of the year. The following is a typical case: X is returned on the schedule as born in September, 1865, and as being 35 years of age. The census, it should be remembered, was taken as of June 1, 1900. Therefore if the date of birth was correct X was not 35 years of age on June 1, 1900, but only 34; if on the other hand he was, in fact, 35 years of age in June, 1900, and was born in September, then the year of his birth was not 1865, but 1864. The most probable explanation of these recurrent inconsistencies is that they result from a rather natural and obvious mistake made in computing the year of birth from the age given. X—or perhaps his wife or some other member of his family, X himself being away from home when the enumerator called — stated that he was 35 years old and was born in September. Then either he or his wife, or more likely the enumerator, subtracted 35 from 1900 and entered the year of birth as 1865. Of course it is possible that the year of birth may have been given directly as 1865, and the year of age then obtained by subtracting 1865 from 1900. This would result in the same inconsistency. To the writer, however, this seems unlikely, because it is his belief and observation that most people keep a better mental record of their age than of their year of birth, and will answer more readily and promptly an inquiry as to age in years than as to year of birth, whether the inquiry applies to themselves or to some relative, friend, or acquaintance.

This conclusion, however, as to the course followed in making the return of age in years does not rest wholly upon inherent probabilities. In the schedules examined, three cases were found where the enumerator had regularly entered the

year of birth with a different ink from that used in making the other entries, thus indicating very conclusively that he had computed and entered the year of birth at the close of his day's field work. Other enumerators may have done the same thing and yet left no trace of this procedure, because they used the same pen and ink that they had been using during the day, or because they copied each schedule entire before sending it in.

One can readily understand why the enumerator resorted to the practice of computing the year of birth if he found that people in general answered the question as to age in years readily enough, but on being asked the year of birth hesitated, became confused, tried to do a sum in subtraction, or went off to hunt up the family Bible. He naturally decided that the easiest way was to ask the age in years and month of birth — most persons know that — and then do his own subtracting in obtaining the year of birth. In the considerable proportion of cases he did it wrong, making the common mistake above described.

When these returns were received at Washington, the Census Bureau was forced to choose between the age return and the year of birth return in cases of disagreement. In view of the fact that the question as to date of birth had been inserted in the schedule on the assumption that it would bring out more accurate replies than the inquiry as to age, thereby reducing the concentration on quinquennial years, the office naturally instructed the clerks who were editing the schedules preparatory to tabulation to follow the year of birth, unless the accompanying data indicated that the year of birth was wrong — and in the rush of getting out the Census reports there was but little time for the slow work of checking by the accompanying data. At any rate, in all those cases where the difference was simply one year, the reported age was almost invariably changed in red ink so as to agree with the reported year of birth, thereby carrying over into the age tabulation the errors which the enumerator—or perhaps the person enumerated—had made in computing the year of birth.

It remains to consider what the effect was upon the age statistics of the Twelfth Census. A reduction of only one year in the ages of one-tenth of the population may not be regarded as a very serious matter, except perhaps as applied to children and young persons. Yet it may account for some things, and one thing which it seems to account for in part is the reduction in the concentration on the multiples of five in the age statistics of the Twelfth Census as compared with those of the Eleventh.

It is undoubtedly true that the enumeration of 1900 was more accurate than that of 1890, just as the enumeration of 1890 was more accurate than that of 1880. The gain in the accuracy of the age distribution of the population shown by the census of 1890 over that of 1880 was due almost entirely to greater care in the enumeration, and the numbers taken from the ages of concentration were distributed quite evenly over the years showing a deficit, as is shown in Table II on page 35. It is a question, however, whether the adjustments made in 1900 in the reported years of age, on the assumption that the reported year of birth was more reliable, did not assign an unduly large proportion of the numbers taken from the ages ending in 5 or 0 to the ages immediately preceding, namely, those ending in 4 or 9.

In order to determine this point the numbers of the total population at the different ages from 23 to 62 inclusive were taken for 1880, 1890, and 1900. A total was made of those at the ages of 23, 33, 43, and 53 at each census and totals also for ages ending in 4, 5, 6, 7, 8, 9, 0, 1 and 2. Percentages were then computed showing the proportion of total population at each age ending with a given integer.

TABLE I.

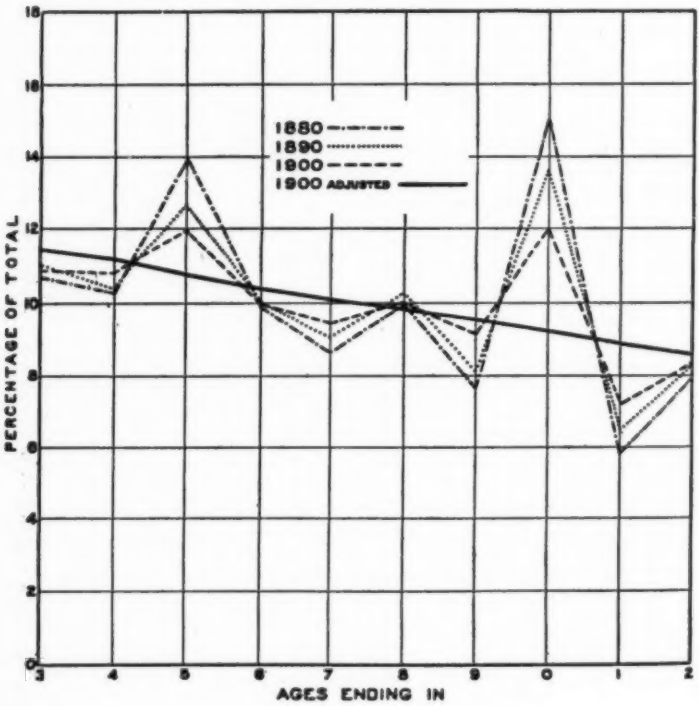
DISTRIBUTION OF TOTAL POPULATION FROM 23 TO 62 YEARS OF AGE: 1880—1900.

Ages ending in	Adjusted ages: 1900*. Per cent.	1900.		1890.		1880.	
		Number.	Per cent.	Number.	Per cent.	Number.	Per cent.
Total.....	100.0	34,144,225	100.0	26,951,141	100.0	20,809,011	100.0
3.....	11.5	3,725,514	10.9	2,963,418	11.0	2,227,913	10.7
4.....	11.2	3,714,557	10.9	2,814,908	10.4	2,155,902	10.4
5.....	10.8	4,088,445	12.0	3,403,799	12.6	2,899,385	13.9
6.....	10.4	3,379,934	9.9	2,711,584	10.1	2,074,816	10.0
7.....	10.1	3,231,259	9.5	2,427,733	9.0	1,800,676	8.7
8.....	9.8	3,426,070	10.0	2,778,471	10.3	2,078,735	10.0
9.....	9.5	3,124,855	9.1	2,176,327	8.1	1,588,307	7.6
0.....	9.2	4,093,618	12.0	3,676,023	13.7	3,127,585	15.0
1.....	8.9	2,498,472	7.3	1,758,842	6.5	1,207,205	5.8
2.....	8.6	2,861,501	8.4	2,240,036	8.3	1,648,487	7.9

*Computed from Appendix A, Supplementary Analysis, Census of 1900, pp. 169 to 173. Here the errors that necessarily creep into census age returns, because of mis-statement of ages, are eliminated, or ironed out, by mathematical interpolation. See Allyn A. Young, "The Adjustment of Census Age Returns," Western Reserve University Bulletin, November, 1902.

From these data the following chart was drawn. This distribution of the adjusted or ideal population, it will be seen, approaches a straight line, while the line of each succeeding census from 1880 to 1900 more closely approximates this ideal. It may be noted in passing that in comparison with the adjusted age distribution, there was a concentration at each census on the ages ending in 5, 8, and 0. The concentration on ages ending in 8 was a very slight one, the years of marked concentration being those ending in 5 and 0.

AGE DISTRIBUTION OF THE POPULATION FROM 23 TO 62 YEARS OF AGE: 1880—1900.



There was a reduction from 1880 to 1890 in the percentage of the population at ages ending in 5 from 13.9 per cent. to 12.6 per cent., or 1.3 per cent. At ages ending in 0 the reduction was from 15.0 per cent. to 13.7 per cent., or 1.3 per cent. There was thus a reduction in the ages of concentration amounting to 1.3 plus 1.3, or 2.6 per cent. There was a similar reduction of 2.3 per cent. between 1890 and 1900.

These reductions are, of course, offset by increases of 2.6 per cent. and 2.3 per cent., respectively, distributed over the remaining ages. The following table shows the distribution of the numbers taken from the ages of concentration.

TABLE II.

Ages ending in	Increase or decrease of per cent.	
	1890 to 1900.	1880 to 1890.
0 and 5	−2.3	−2.6
1 and 6	0.6	0.8
2 and 7	0.6	0.7
3 and 8	−0.4	0.6
4 and 9	1.5	0.5

The distribution of the numbers taken from the ages of concentration was much more regular in the census of 1890 than in that of 1900. Nearly three-fourths of the amount taken from the ages ending in 5 and 0 in 1900 was credited to ages ending in 4 and 9. While the concentration on ages ending in 5 and 0 was thus considerably reduced, the reduction was accompanied by a biased distribution of the numbers transferred to other ages. It was not to be expected, of course, that this distribution would be absolutely regular. As the general tendency of people is to understate ages,* any concentration on years ending in 0 and 5 is drawn largely from the years immediately succeeding, i. e., those ending in 1 and 6; and any increase in the accuracy of age returns will,

*See Supplementary Analysis, Census of 1900, p. 130.

in lessening the concentration on years ending in 0 and 5, add more to the years ending in 1 and 6 than other years. This was true in 1890, as Table II shows. In 1900, also, the increase for the years ending in 1 and 6 was a comparatively large one, although not half so large as the increase for the years ending in 4 and 9. The latter increase was undoubtedly brought about by means of the clerical adjustments considered above, which consisted largely in reducing the reported age one year, to conform to the reported date of birth.

The problem was approached from still another angle. It has already been stated that a careful examination of 130,000 representative urban and rural census enumerations of 1900 disclosed lack of agreement between reported age and date of birth in nearly one-tenth of the age returns. There were 12,526 such cases of disagreement, which it was necessary to adjust in some manner. Of the 12,526 clerical adjustments made 8,851, or 70.7 per cent., reduced the reported ages by one year and 950,† or 7.6 per cent., by one month, while the remainder comprised all manner of emendations.

The 8,851 corrections representing reductions of one year were found to be of the following ten types.

TABLE III.

Reductions		Number.
From ages ending in	To ages ending in	
0	9	1,150
1	0	588
2	1	969
3	2	851
4	3	866
5	4	1,102
6	5	893
7	6	856
8	7	882
9	8	694

†These cases were children less than one year old, whose ages were reported in terms of months.

It will be seen that the ages ending in 0 lost 1,150 to the ages ending in 9, and received 588 from the ages ending in 1; the net loss on the part of the ages ending in 0 was therefore 562. Similarly, the ages ending in 5 suffered a net loss of 209. For ages ending in 0 and 5 the net loss, then, was 562 plus 209, or 771. The net gain or loss for all the age groups is shown in the following table:

TABLE IV.

Ages ending in	Net gain or loss
0 and 5	−771
1 and 6	+344
2 and 7	− 92
3 and 8	−173
4 and 9	+692

Thus the net loss from the ages ending in 0 and 5 is nearly balanced by the net gain to the ages ending in 4 and 9; in other words, the one-year adjustments did, it is true, reduce the concentration on multiples of 5, but loaded that reduction almost entirely upon the ages ending in 4 and 9, where only part of it properly belonged. The net increase in the ages ending in 1 and 6 was practically balanced by the reductions in the remaining ages.

The excess per cent. of concentration on multiples of 5 was estimated by Dr. Young as follows:*

Census of 1880	44.8 per cent.
Census of 1890	31.3 per cent.
Census of 1900	19.8 per cent.

If the relative number of one-year reductions in the 130,000 enumerations examined was typical of all the census returns for 1900,—and the schedules examined were representative of urban and rural conditions in all sections of the country,—the reductions made for the whole United States must have

*Supplementary Analysis, Census of 1900, p. 137.

been responsible for a decrease of 2.9 per cent.† in the concentration on multiples of 5. If the reductions had not been made, therefore, the excess concentration in 1900 would have been 22.7 per cent., and not 19.8 per cent.

Looking at the question from every point of view, then, it seems clear that in the ages as actually returned on the Twelfth Census schedules the decrease of concentration was not so great as it appeared to be after these returns had been adjusted to agree with the return of the year of birth; and since the return of the year of birth was more inaccurate than the return of the age in years, it follows that the gain in accuracy of the age statistics of the Twelfth Census was somewhat smaller than has been supposed, or was, in other words, to some extent a delusion.

As the actual distribution of the population by age groups comes in succeeding census enumerations to correspond more and more nearly to that of an adjusted or ideal population, the gain in accuracy will become gradually smaller. The concentration on certain ages will probably never disappear, as it is a function of two variables: the care of the enumerator in asking the questions and recording the answers, and the ignorance or insincerity of the person questioned. It is desirable that each census should show improvement over its predecessors, but it seems probable that the improvement in the census of 1900 was not so great as has been believed. From the facts adduced in the preceding pages the following conclusions were drawn:

(1) That the inclusion in the census schedule in 1900 of an inquiry as to date of birth was rendered valueless in a considerable proportion of the cases by the practice of enumerators, or others, of computing date of birth from the age given.

(2) That in a considerable proportion—perhaps one tenth —of the cases the year of birth was incorrectly computed, the

†This figure was obtained as follows: The number of ages ending in 0 and 5 is normally one fifth of all the ages in any given age series. The normal number of such ages in the 130,000 cases under consideration would thus be 26,000. The ratio of 771, the net loss on the part of such ages, as shown in Table IV, to 26,000, is 2.9 per cent.

extent of the inaccuracy being in a large majority of instances exactly one year.

(3) That the assumption on the basis of which the schedules were corrected in the office of the census, i. e., that where the date of birth and the age in years were in disagreement, the former was the more reliable, is not borne out by a careful study of the internal evidence in the schedules themselves.

(4) That the adjustments made on this assumption resulted in an extensive reduction of ages by one year, which by transferring a large number of the ages ending in 5 and 0 to the ages ending in 4 and 9, fostered the belief that the improvement in the age returns of 1900 was somewhat greater than was probably the actual case.

In view of the conclusions just presented, and especially in view of the first conclusion, that the inquiry as to date of birth played little or no part, either in increasing the accuracy of the age returns of the Twelfth Census, or in reducing the concentration on years ending in the integers 5 and 0, it was decided to eliminate the query regarding date of birth from the population schedule of the Thirteenth Census, and to retain only the question regarding age at last birthday.

THE SOCIAL MARKING SYSTEM.*

By Frankin H. Giddings, *Professor of Sociology, Columbia University.*

In their quantitative aspect many data of sociology are positions in a scheme, rather than distances from zero. Tables of births, deaths, and migrations, like tables of height and of weight, give us measures from zero. Degrees of difference or of resemblance which we observe among our acquaintances, and others, the affiliations of nationalities and of races, of religious belief, and of political interests, the values that we assign to ability and to conduct, and the social ranks that make up differentiated communities, are merely positions in a scheme.

Sir Francis Galton long ago showed that when positions in a scheme are successive, and may be successively numbered, and when the instances in which given phenomena fall into successively numbered positions, can be counted, and their frequencies set down, we can legitimately subject the numerical data so assembled to familiar methods of statistical analysis. We can plot their rough curves of magnitude and of frequency; we can ascertain their medians, quartiles, and probable errors; we can determine their modes. Whether we can also make significant use of their standard deviations and coefficients of variation, is a question that need not now be considered.

In order to assemble and to sort such data, however, we must have a marking-scale. Also we must have a concrete knowledge of the facts to be sorted, like that on which we rely when we assign numerical grades to examination papers, or to tests of conduct. The marking-scale must not be arbitrary, a mere ingenious invention. It must be a simple and natural expression of observed relations. If possible, its elements

*Address delivered at a joint meeting of the American Statistical Association and the American Sociological Society, New York, December 29, 1909.

should be derived from those common-sense gradings which the world uses in rough and ready fashion for the purposes of everyday life, and which therefore have the warrant of experience.

A familiar example of empirical gradings is the scale of consanguinity. The children of brothers or sisters we call first cousins, and describe them as one degree less nearly related than brothers or sisters. Children of first cousins we call second cousins, and describe them as two degrees less nearly related than brothers or sisters, and so on. These degrees, as everybody recognizes, are nothing more than successive positions in a scheme. No absolute numerical value, or distance from zero, attaches to any one of them. A kindred but less familiar example is offered in Lewis H. Morgan's ordinal arrangement of the successively wider and wider groupings composing tribal society. The arrangement is, namely: first grouping, the family; second grouping, the totem kin or clan; third grouping, the phratry or brotherhood of clans; fourth grouping, the tribe; fifth grouping, the federation of tribes.

One intent of this paper is to call attention to other rough marking-scales which we daily make use of in describing social phenomena, and to show how they may be extended and made sufficiently precise for scientific purposes. A second intent is to indicate how, by distributing a considerable mass of existing numerical data in accordance with the marking-scales to be described, we may, through the usual methods of statistical analysis, arrive at sociological conclusions that are interesting and possibly important.

To a great extent the phenomena of population are described by statistical measures from zero. Density, excess of births over deaths, emigration and immigration, are expressed in absolute numbers, of approximate accuracy. But such measures are not forthcoming when we ask a question like the following, namely, Is a population made up of native white Americans, foreign-born Irishmen and Germans, foreign-born Italians, Hungarians and Poles, and both native and foreign negroes, more mixed, composite, or heterogeneous than one made up of native white Americans, Spanish-Ameri-

cans, Chinamen, Hawaiian Islanders, and Indians not taxed, and, if so, how much more? To answer a question of this kind we must first assign each nationality or race to a position on a predetermined marking-scale of resemblance or affiliation. We must be able to say: Nationality E is so many points more unlike the native-born whites than Nationality B is, or to say: Race K is psychologically and culturally so many positions more distant from Race A than Race D is.

By general consent we have employed a scale of this kind for twenty years or more in our discussion of the immigration problem. No one ever questions the propriety of assigning the native-born of white parents to the initial position, which we will call "zero," nor of assigning the native-born of foreign-born whites to the second position which we will call "one." In like manner, without dispute the uncivilized races are assigned to the final position, whatever its number may happen to be. Assignments to intermediate positions are made by concurring judgments upon the varying degrees of resemblance which immigrant nationalities bear to the native-born whites of native parentage. Immigrants from the British Isles and English provinces of Canada, who have spoken the English language from birth, are held to be more like the native stock than are the immigrants from continental Europe. Of the latter, we regard the people of northwestern Europe as psychologically, culturally, and historically more nearly related to us than are the peoples of southern or of eastern Europe.

But little interpolation and rearrangement are needed to convert this rough marking-scale of nationalities and races into a scientific scale of ten positions, as follows: 0, Native born of native white parents; 1, Native-born of foreign-born white parents; 2, Foreign-born, English-speaking; 3, Northwestern Europeans; 4, Southern Europeans and Latin-American whites; 5, Eastern Europeans; 6, All other whites; 7, Civilized yellow; 8, Civilized dark; 9, Uncivilized.

The principle on which this marking-scale is constructed for the population of the United States may be applied to the population of any country by observing a general rule: As-

sign to the third position (numbered 2), all foreign-born elements speaking the language of the country to which they have come. Assign to the fourth position (numbered 3) the foreign-born that are psychologically, culturally, and historically more nearly related to the nationalities of position three than any remaining element is. Assign to position five (numbered 4) the foreign-born that are psychologically, culturally, and historically more nearly related to the nationalities of position four than any remaining element is, and so proceed.

In deciding what nationalities are "more nearly related" in the ways specified, we must not substitute the judgment of an individual observer for the concensus of popular opinion. In matters of this kind the "probable error" of thousands of individual judgments or markings making up the common opinion is extremely small.

There is, however, a scientific way of checking or testing the popular marking. It consists in breaking up each mark into a number of units, each of which stands for some specific and undisputed objective fact.

We may apply this test to the foregoing scale of ethnic positions, as follows:

Build up a composite statistical mark for "Americanism" by assigning one unit to each of the following facts, namely: (1) Parents native-born; (2) Self native-born; (3) Native language, that of the United States (English); (4) Reared under Celto-Teutonic traditions and cultural influences; (5) Reared under constitutional government; (6) Of European stock and reared under European civilization; (7) Belongs to race that has created an independent political state with a history; (8) Belongs to race that has created an ethical religion; (9) Belongs to race that has created a literature; (10) Belongs to race that has independently risen above barbarism; (11) Color lighter than yellow; (12) Color lighter than red; (13) Color lighter than brown; (14) Color lighter than black.

Assigning these units of marking to the respective elements composing the population of the United States, we get the following grades or ratings:

MARKINGS SCHEME FOR NATIONALITIES.

	Parents Native Born.	Self Native Born.	Native Language English.	Reared under Celtonic Teutonic Traditions and Culture.	Reared under Constitutional Government.	Of European Stock and Reared under European Civilization.	Belongs to Race that Has Created an Independent Political State.	Belongs to Race that Has Created an Ethical Religion.	Belongs to Race that Has Created a Literature.	Belongs to Race that Has Independently Risen above Barbarism.	Color Lighter than Yellow.	Color Lighter than Red.	Color Lighter than Brown.	Color Lighter than Black.	Total Mark.	Position.
Native born of native parents...	1	1	1	1	1	1	1	1	1	1	1	1	1	1	14	0
Native born of foreign parents..	:	1	1	1	1	1	1	1	1	1	1	1	1	1	13	1
Foreign born English speaking.	:	:	1	1	1	1	1	1	1	1	1	1	1	1	12	2
Northwestern European.....	:	:	:	1	1	1	1	1	1	1	1	1	1	1	11	3
Southern European.....	:	:	:	:	1	1	1	1	1	1	1	1	1	1	10	4
Eastern European.....	:	:	:	:	:	1	1	1	1	1	1	1	1	1	9	5
All other whites.....	:	:	:	:	:	:	1	1	1	1	1	1	1	1	8	6
Asian yellow.....	:	:	:	:	:	:	1	1	1	1	:	1	1	1	7	7
Civilized dark	1	1	1	1	1	:	:	:	:	:	:	:	:	:	5	8
Uncivilized	:	:	:	:	:	:	:	:	:	:	:	:	1	1	2	9

Passing from phenomena of population to phenomena of the social mind, we observe that simple marking-scales are daily used in connection with religion, conduct, ability, and education. The empirical gradings applied to these phenomena appear on examination to be not less natural or matter-of-course than the popular gradings of nationalities in respect of their resemblances.

If a nation has an established religion, it is assigned to the initial position. All other religions are looked upon as variate, kindred, or alien cults. In the United States we have no established religion, and we therefore assign to the initial position the denominations that in the earlier days of our history were proclaimed by a majority of the people to be orthodox, notwithstanding the fact that they were Protestant bodies, holding radical ideas on questions of church government. These denominations, after a time, divided on questions of doctrine, some of them becoming radical in belief, as well as in organization. Protestant bodies conservative in both belief and organization were formed at an early period, but for a time they were commonly regarded with disfavor. It was much later that the Roman Catholics gained a foothold outside of Maryland, and only recently have come the Eastern Catholics and the non-Christian persuasions.

The essential relationships of these varying cults present a scheme or order closely correlated with the historical succession. The conservative Protestants, for example, are in matters of faith and worship more nearly related to the radical Protestants, than the Roman Catholics are. The Roman Catholics, in turn, are more nearly related to the conservative Protestants than the Eastern Catholics or the non-Christian bodies are.

Accepting then the popular judgment upon these points of relationship, our marking-scale is as follows: Initial position or 0, Protestant radical A, including all religious bodies that are radical in both doctrine and organization; 1, Protestant radical B, including all Protestant bodies that are radical in organization only; 2, Protestant conservative; 3, Catholic A, including all Roman Catholics; 4, Catholic B, including all

Eastern Catholics; 5, Nominally Christian, but in many particulars reverting to the patriarchal Jewish type, namely, the Latter-Day Saints, or Mormons; 6, Non-Christian monotheistic, principally Jews; 7, All others.

This scale, expressive of popular judgment, may be tested, as we tested the scale for nationalities, by assigning unit marks to objective facts, one at least of which can be affirmed of all the religious groups, and all of which can be affirmed of at least one group. It will be found that the following facts meet these conditions, and we therefore assign a unit mark to each one: (1) Radical in doctrine; (2) Radical in government; (3) Protestant; (4) Holding to Western European Christian tradition; (5) Holding to general European Christian tradition; (6) Holding to generic Christian tradition; (7) Monotheistic; (8) Theistic.

Assigning these units of marking to the various groups of religious bodies found in the United States, we get the following ratings:

MARKING SCHEME FOR RELIGIONS.

	Radical in Doctrine.	Radical in Government.	Protestant.	Holding to Western European Christian Tradition.	Holding to General European Christian Tradition.	Holding to Generic Christian Tradition.	Monotheistic.	Theistic.	Total mark.	Position.
Protestant, radical, A	1	1	1	1	1	1	1	1	8	0
Protestant, radical, B	..	1	1	1	1	..	1	1	7	1
Protestant, conservative	..	..	1	1	1	..	1	1	6	2
Catholic, A	..	..	..	1	1	1	1	1	5	3
Catholic, B	..	..	..	..	1	1	1	1	4	4
Mormon	..	..	..	..	..	1	1	1	3	5
Jewish	..	..	..	..	..	..	1	1	2	6
All others	..	..	..	..	..	..	..	1	1	7

The marking-scale applied to conduct is an inheritance from the earliest days of civilization, and is substantially the same throughout the civilized world. Its gradations designate objectionable, rather than commendable, qualities of behavior. To the initial position 0, we assign the inoffensive and law-abiding; to 1, we assign the minor misdemeanants; to 2, the major misdemeanants; to 3, the felons whose offenses are not capital; to 4, the felons whose offenses are capital.

On the marking-scale for ability, we assign to the initial position those who are competent to work under direction; to 1, we assign those who are capable of self-direction in a small way; to 2, those who can supervise in accordance with general instructions provided for them; to 3, those competent to exercise responsible direction, a group which includes the great majority of land-owning farmers, business men, and professional men; and to 4, we assign the directors of directors, namely, the greater organizers, statesmen and financiers, the greater discoverers and inventors, and the greater professional men and philosophers.

The marking-scale for education, in like manner, is simple, and is the same for all occidental countries.

The groups or classes assigned to its successive positions are: 0, the illiterate; 1, those educated not beyond the grammar-school grades; 2, those educated not beyond the high-school grades; 3, those educated not beyond the college grades; 4, those educated in university grades.

Assignments to positions on the conduct-scale and education-scale are matters of record, and usually it is not necessary to go behind the returns. Each grade, however, and each grade of ability admits of indefinite subdivision, and assignments to the intermediate positions must be made by experts. The probable error of their gradings is the measure of the scientific value of the marks. The most admirable example of the possibilities of such gradings that we have is found in Professor Cattell's *Directory of American Men of Science.*

We may pass now to the inquiry: What sociological problems may be stated and solved in statistical terms with the aid of these social marking-scales?

The answer turns upon a technical point, which may be stated in the form of a preliminary question, namely, Are the "distances" or "steps" between successive positions on any given marking-scale always equal, or may they be regarded as equal for the purposes of sociological induction? May we assume, for example, that the step from "native-born of native parents" to "native-born of foreign parents" is neither longer nor shorter than the step from "native-born of foreign parents" to "foreign-born, English-speaking?" Or may we assume that the step from "Protestant radical B" to "Protestant conservative" is equal to the step from "Protestant conservative" to "Catholic A?" If these steps or any others on the same scale are unequal, we may still find the scale worth while, and from the data to which it is applied we may draw conclusions within those limits that have been indicated by Galton, Bowley and Thorndike. If, however, the steps are equal, or may legitimately be treated as if they were equal, the use of the scale is greatly extended.

What then are the facts? Is it possible for any individual or group of experts to prove, or for the general public to "feel sure," that the "steps" in these various marking-scales are distinguishable in magnitude? If it is alleged that they are distinguishable, and if the differences can be specified, we can overcome the difficulty, at least in part, by weighting our marks. If, on the other hand, it is admitted that no one can say whether the "steps" are equal, or not, a fact of great consequence to scientific sociology is thereby established.

The significance of differences and inequalities among factors of a social population lies in the fact that they are reflected in social feeling and impair the sense of social solidarity, thereby affecting co-operative efficiency. Race conflicts, religious wars, and persecutions, and class conflicts, are expressions of a conscious antagonism which depends for its existence upon a consciousness of difference. If there is no feeling of difference, there is no prejudice; and if differences are too slight or too vague to be detected, the feeling of difference cannot arise.

Therefore, differences among social factors which are not

perceived or felt by the social units themselves are without effect upon the important phenomena of prejudice, the consciousness of kind in general, and the sense of social solidarity. For the purposes of sociological theory they may be neglected.

Similar reasoning applies to inequalities between any two differences and any other two. The difference between one nationality and another may be obvious, but it may be impossible to perceive or even to feel that the difference between nationalities C and D is greater than the difference between nationalities A and B. This is precisely the kind of difference that we contemplate when we ask whether the "steps" in any given social marking-scale are equal.

We conclude, then, that when a social marking-scale has been so constructed that it is not possible for the expert or the public to say whether the successive "steps" of the scale are equal, or not, any absolute difference which may undetected exist is negligible for the purposes of sociological theory. The "steps" may be assumed to be equal.

Adopting this conclusion, we proceed to show how the social marking-scales may be used in the statistical analysis of sociological problems.

The following table presents the census data on population for the years 1890 and 1900, grouped according to the marking-scale for nationalities and races (Table I):*

* I am indebted to Mr. F. Stuart Chapin for the computations involved in preparing the tables for this paper.

TABLE I.

RESEMBLANCE POSITONS OF NATIONALITIES AND RACES: CONTINENTAL UNITED STATES.

Nationalities and Races.	Positions.	1890.		1900.	
		Number.	Percentage.	Number.	Percentage.
Native born of native white parents	0	34,468,602	.547	41,356,097	.541
Native born of foreign white parents	1	11,503,675	.182	15,687,322	.205
Foreign born, English speaking	2	3,807,337	.060	3,581,113	.047
English		908,141		842,078	
English Canadian		678,442		785,958	
Scotch		242,231		233,977	
Welsh		100,079		93,682	
Irish		1,871,509		1,618,567	
Australian		5,984		6,851	
Not specified		951		.....	
Foreign born, Northwestern Europe	3	4,468,502	.071	4,911,163	.064
French		113,174		104,341	
French Canadians		302,496		395,297	
Belgians		22,639		29,804	
Germans		2,784,894		2,668,990	
From Luxemburg		2,882		3,041	
German Poles		.....		150,232	
Hollanders		81,828		105,049	
Danes		132,543		154,284	
Swedes		478,041		573,040	
Norwegians		322,665		336,985	
Swiss		104,069		115,851	
Austrians		123,271		276,249	
Southern Europeans and Latin Americans	4	313,955	.005	668,119	.009
Spaniards		6,185		7,072	
Portuguese		15,996		30,618	
Italians		182,580		484,207	
Greeks		1,887		8,564	
Mexicans		77,853		103,410	
Central Americans		1,192		3,901	
South Americans		5,006		4,761	
Cubans		.....		11,153	
West Indians		23,256		14,433	
Eastern Europeans	5	510,623	.008	1,038,019	.014
Russians		182,644		424,096	
Finns		.....		62,811	
Russian Poles		.....		154,424	
Austrian Poles		.....		58,503	
Poles		147,440		.....	
Unknown Poles		.....		20,351	
Hungarians		62,433		145,802	
Bohemians		118,106		156,991	
Roumanians		.....		15,041	

TABLE I.—*Concluded.*

Nationalities and Races.	Positions.	1890.		1900.	
		Number.	Percentage.	Number.	Percentage.
All other whites*	6	32,376	.0005	35,319	.0004
Turks		1,839		9,933	
Europeans not specified		12,579		2,263	
Born at sea		5,533		8,229	
Other countries		479		2,558	
Atlantic Islands		9,539		9,784	
Africa		2,207		2,552	
Civilized Yellow	7	111,240	.002	118,812	.002
Chinese		106,688		81,827	
Japanese		2,292		25,077	
Asian (not specified)		2,260		11,908	
Civilized Dark	8	7,553,794	.120	8,869,342	.116
Civilized Negroes		7,488,676		8,749,735	
Civilized Indians		58,806		107,706	
Hindoos		2,143		2,050	
Sandwich Islanders		1,304		.....	
Pacific Islanders		2,065		2,049	
Uncivilized and otherwise not specified	9	189,447	.003	129,518	.002
Uncivilized Indians		189,447		129,518	
Totals		62,959,551	.100	76,394,824	.100
Sub-homogeneity		1.579		1.552	
Sub-homogeneity from percentages		.01579		.01552	

Plotting the data of this table as a surface of magnitude, according to Galton's method,† we get the rough curves of diminishing magnitude, or sub-homogeneity, shown in Figs. I and II. Plotting as a surface of frequency, we get the rough curves of frequency shown in Figs. III and IV.

*The group "All Other Whites" as here made up should in strict accuracy be distributed. It should include Turks, Persians, Armenians, African whites, not descended from European stock, etc., while "Europeans not Specified," "Born at Sea," and from "Atlantic Islands," should be assigned to preceding groups. The labor of ascertaining what those assignments should be was not worth while, the entire group constituting only .0004 per cent. of our total population.

† *Natural Inheritance.* pp. 37, 38. Galton uses the terms "polygon of distribution" and "curve of distribution" to distinguish from "polygon of frequency" and "curve of frequency," the forms obtained by plotting magnitudes on the vertical y, instead of on the horizontal x. Inasmuch as the "curve of frequency" is also a "curve of distribution," it is more exact to use the term "surface orcurve of magnitude" when we plot magnitudes on the vertical and frequencies on the horizontal, and the term "surface or curve of frenquency" when we plot magnitudes on the horizontal and frequencies on the vertical.

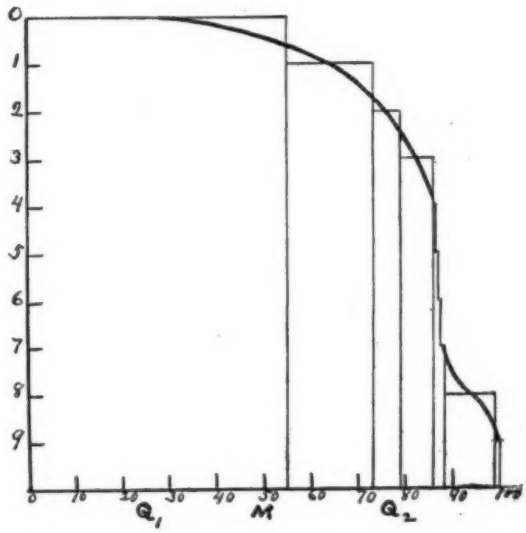


Fig. 1. Nationalities: Magnitude curve of resemblance positions, 1890.

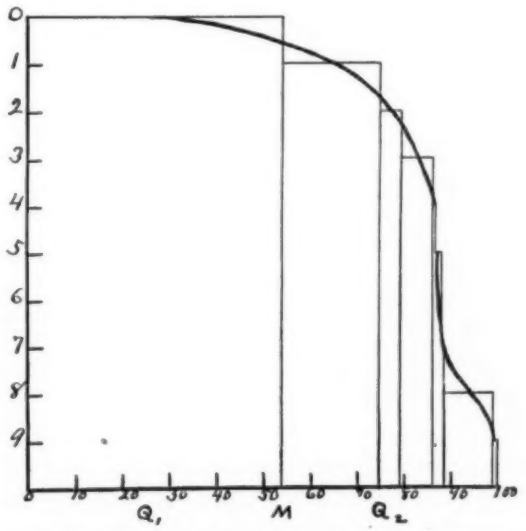


Fig. 2. Nationalities: Magnitude curve of resemblance positions, 1900.

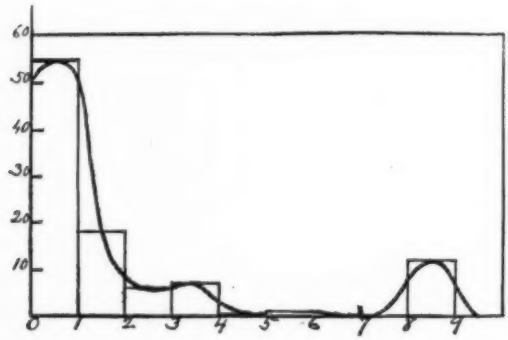


Fig. 3. Nationalities: Frequency curve of resemblance positions, 1890.

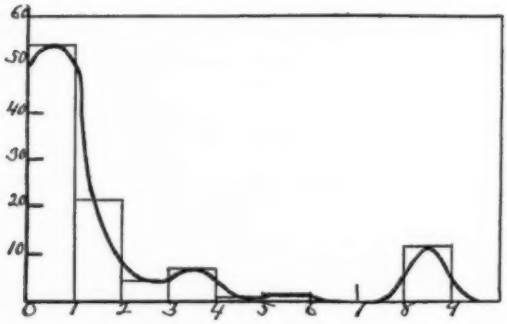


Fig. 4. Nationalities: Frequency curve of resemblance positions, 1900.

These curves show in a rather striking way the more significant relations of our numerous ethnic elements to one another and to the total population, and in particular the extent of departure from an ideal homogeneity. The true measures of this departure are the positions of the medians and the quartiles. Notwithstanding the magnitude of the immigration which we have received within the last twenty-five years, nearly three-fourths of the entire population of continental United States is native-born, and more than one half

is native-born of native parents. Nearly 80 per cent. has been English-speaking from birth. These facts of course are disclosed by the census data independently of the marking-scale here used. But the scale and the plotted curves reveal one tendency that, without their aid, would not easily be discovered, namely, the curve of 1900 is slightly less irregular than the curve of 1890. Very slightly, but perceptibly, it is nearer the straight line which would express the divergence of nationalities from the native-born of native parents, if each successive step of divergence were equal to the step preceding, and if each nationality were represented in our population by the same number of individuals. This circumstance suggests a question: Is our population becoming slightly less heterogeneous contrary to the prevailing belief? We shall presently see that it is.

The curves of frequency show the positions of the modes. A multi-modal curve, as all statisticians understand, indicates a compound group, or one made up of two or more kinds of units. The American population is racially compound, and the actual gulf between the native-born whites and the negro element is clearly shown by the primary mode at 0 (Figs. III and IV), and the secondary mode at 8. Moreover, a tendency toward the formation of lesser modes among the foreign-born, slightly manifest in 1890, is distinctly more marked in 1900.

Table II presents the census data on religious bodies for the years 1890 and 1906, grouped according to the marking-scale for religion.

Plotting the data of this table as surfaces of magnitude, we get the rough curves of sub-homogeneity shown in Figs. V. and VI. Plotting as surfaces of frequency we get the rough curves of frequency shown in Fig. VII and VIII.

TABLE II.

RESEMBLANCE POSITIONS OF RELIGIOUS BODIES: CONTINENTAL UNITED STATES.

Religious Bodies.	Position.	1890.		1906.	
Protestant, radical, A	0		1.3		1.3
Adventist		0.3		0.3	
Christians		0.5		0.6	
Christian Science		0.0		0.3	
Unitarians		0.3		0.2	
Universalist		0.2		0.2	
Protestant, radical, B	1		51.8		45.7
Baptist		18.0		17.0	
Congregationalist		2.5		2.1	
Disciples		3.1		3.5	
Dunkers		0.4		0.3	
Evangelical bodies		0.6		0.5	
Friends		0.5		0.3	
German Evangelical		0.9		0.9	
Independent		0.1		0.2	
Mennonites		0.2		0.2	
Methodist		22.3		17.5	
Reformed		1.5		1.4	
United Brethren		1.1		0.9	
Other Protestant		0 6		0.7	
Protestant, conservative	2		14.8		14.7
Lutheran		6.0		6.4	
Presbyterian		6.2		5.6	
Protestant Episcopalian		2.6		2.7	
Catholic, A	3		30.3		36.7
Roman Catholic		30.3		36.7	
Catholic, B	4		0.0		0.4
Eastern Orthodox		0.0		0.4	
Mormon Christian	5		0.8		0.8
Latter-Day Saints		0.8		0.8	
Non-Christian, A	6		0.6		0.3
Jewish		0.6		0.3	
Non-Christian, B	7		0.3		0.2
All non-Christian religious bodies not included under non-Christian, A		0.3		0.2	
Sub-homogeneity from percentages		1.833		1.953	

If marking-scales of social phenomena, and the curves of magnitude and of frequency which we are able to plot by means of them, had no other use than that of enabling us to form approximately correct notions about the composition of a population, and the make-up of its social mind, they would be serviceable enough to repay the labor involved in applying them. They have, however, a further and more important use, which must now be indicated.

The question has been raised whether the American popula-

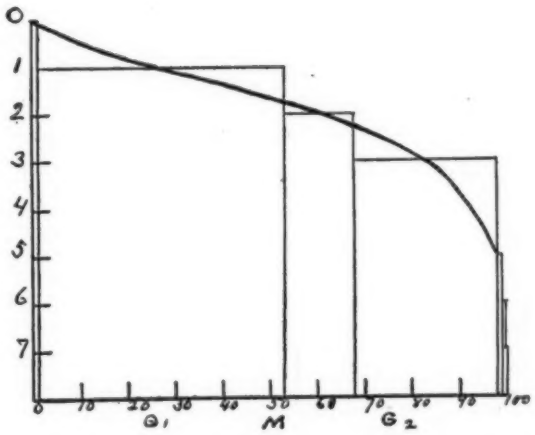


Fig. 5. Religious: Magnitude curve of resemblance positions, 1890.

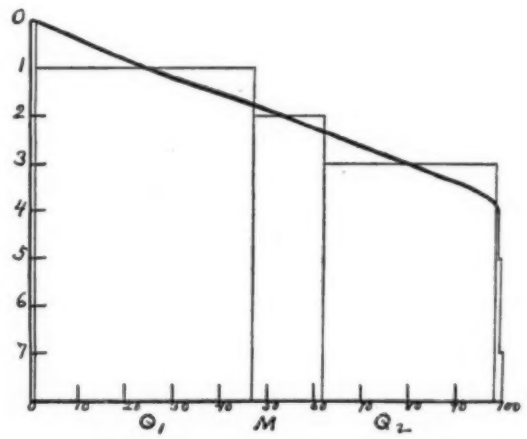


Fig. 6. Religious: Magnitude curve of resemblance positions, 1906.

tion is becoming more or less heterogeneous in ethnic composition. Not less interesting are the questions: Are we becoming more heterogeneous or less heterogeneous in religious belief?

Are we becoming more unlike or less unlike in conduct? Are we becoming more unequal or less unequal in educational attainments?

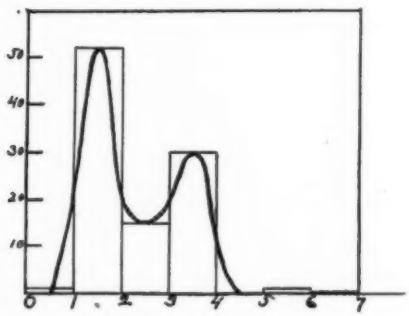


Fig. 7. Religious: Frequency curve of resemblance positions, 1890.

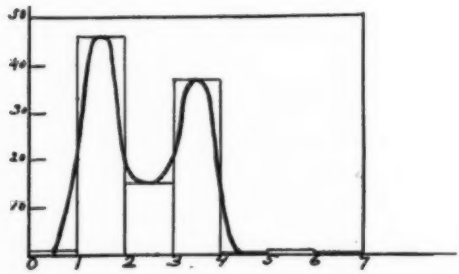


Fig. 8. Religious: Frequency curve of resemblance positions, 1906.

But to ask these questions is to raise yet another which is preliminary to them all, namely: Is it in any case possible to measure relative heterogeneity? Given two compounds, is it possible to determine how much more compounded one is than the other?

I do not know what attempts may have been made to answer this question in its generalized form, but I believe that no method has hitherto been suggested, or at any rate used, for measuring the sub-homogeneity of mixed populations, or of the social mind. I venture therefore to suggest that the problem admits of solution and to offer a formula for finding a numerical measure or coefficient of heterogeneity, or, as I think we may more accurately say, of sub-homogeneity.

When we plot the surface of magnitude (the polygon of distribution as defined by Galton), we measure magnitudes, or we determine successive positions, by vertical distances from a baseline. Since equal magnitudes must be measured, and positions that indicate equal degrees of relationship must be determined, by equal vertical distances, the plotted points indicating such equal measures must lie in a horizontal straight line. Therefore, the horizontal straight line is the graphic expression of equality or of homogeneity. Departures from equality, or degrees of sub-homogeneity, are best plotted as minus or negative magnitudes, graphically expressed by points placed at proper distances below a horizontal line, which itself expresses the ideal homogeneity from which departures are supposed to be made.

Let us then suppose that we have to describe a group of human beings, twelve in number, of whom three are unobjectionable in conduct; three are vicious persons, or minor misdemeanants; two are petty criminals or major misdemeanants; two are felons, not capital; and two are felons, capital. The group as a whole is sub-homogeneous. Plotting the data, we get the arrangement shown in Fig. IX.

Imagine now, that by expending one unit of some kind of effort, we could lift any one vicious person up to the level of the men whose conduct is unobjectionable; that by expending two units of the same kind of effort we could lift any one major misdemeanant to the same standard level; that by expending three units of the same kind of effort, we could lift any one of the minor felons, and by expending four units of the same kind of effort we could lift any one of the major

felons, to our standard level. Then, by expending 1 unit $\times 3 + 2$ units $\times 2 + 3$ units $\times 2 + 4$ units $\times 2$, or 21 units in all, or 1.75 units per capita, for the whole group of twelve persons, we should convert the entire sub-homogeneous group of twelve persons into a group perfectly homogeneous in respect of a standardized conduct.

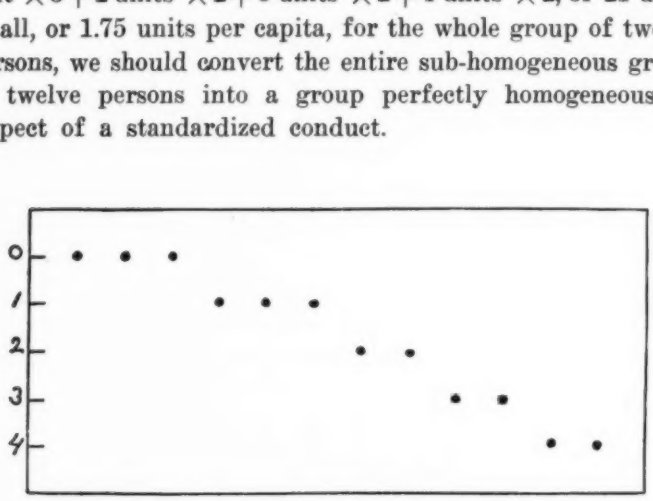


Fig. 9. Measurement of Sub-Homogeneity.

Students of the physical sciences who are accustomed to measure physical phenomena of every description by the number of units of effort, or of energy, necessary to transform them from one state into another, will assent to the proposition that if we could thus actually transform any heterogeneous group of human beings into a homogeneous group, the number of units of effort necessarily expended in the process could be taken as an accurate measure of the total sub-homogeneity of the original group, and that the total number of units of effort so expended, divided by the total number of individuals in the transformed group, could be taken as a measure of the per-capita degree of sub-homogeneity of the original group.

Is it legitimate to conceive of any heterogeneous group as ideally transformable by such a procedure, and then to assume that we may measure its sub-homogeneity by (1) multiplying each successive numerical mark on a marking-scale of resemblance positions by the "frequency" or number of individuals

assigned to that position, (2) obtaining the sum of the products, and (3) dividing it by the whole number of individuals in the group or population?

It is, I think, an adequate and satisfactory answer to this question to observe that the conception and the assumption are legitimate, if mankind is warranted in believing that by an expenditure of educational and reformatory effort, it can standardize knowledge and conduct, and can assimilate alien habits and ideals to prevailing or national types. If the validity of this pragmatic belief be conceded, there can be no objection to conceiving of an average and abstract unit of standardizing effort, practically unchanging throughout the same group or population, living under practically constant conditions.

If so much be granted, we may write the formula for measuring sub-homogeneity as follows:

Designate positions on the marking-scale by the numerals, 0, 1, 2, 3, 4, ... n.
Designate frequencies by K_0, K_1, K_2, K_3, K_4, ... Kn.
Designate total individuals, or Population by P.
Designate per capita degree of sub-homogeneity by S.

Then:

$$S = \frac{K_1 + 2K_2 + 3K_3 + \ . \ . \ . \ + NKn}{P}$$

Applying this formula to the data presented in the tables of ethnic distribution by resemblance position, we obtain for 1890 the coefficient .01579 as our expression for the per-capita degree of sub-homogeneity, and for 1900 the coefficient .01552. Applying the formula to the data presented in the tables of religious distribution by resemblance positions, we obtain for 1890 the coefficient 1.833, and for 1906 the coefficient 1.953. These coefficients indicate that in ethnic composition the people of the United States are now becoming, contrary to the prevailing impression, slightly more homogeneous, but that in religious persuasion they are still becoming slightly more heterogeneous.

THE NECESSITY OF THE SUPERVISION OF WEIGHTS AND MEASURES.

By Fritz Reichmann, *State Superintendent of Weights and Measures, Albany, N. Y.*

Statistical results and deductions therefrom are based upon the assumption of the uniformity of the units, the uniformity of the standards of measurement. When these deductions or results are of a very general character and no close degree of accuracy is intended this assumption may be safe. There are many phases of commercial activity in which there is little or no uniformity of standards and there exists a greater degree of heterogeneity than was formerly true in physical or chemical work where it was often difficult to interpret or repeat certain investigations because of the lack of knowledge of what the standards were.

In the ultimate analysis, practically every transaction of barter or trade is based upon some standard of quantity. How many cubic feet of gas? How many kilowatt hours of electricity? How many miner's inches of water? How many tons of coal? How many acres of land? How many bushels of certain farm produce? All of these questions must be answered in terms of quantity. It would seem, therefore, highly important to have definite, uniform, reliable, consistent standards by which transactions, wholesale or retail, are gauged. This is particularly true in a democratic form of government where there is still in the minds of many a confusion of liberty and license. There are many, even some in exalted position, who decry any form of necessary restraint. The framers of the constitution of the United States when defining the powers of Congress very properly provided for the establishment and enforcement of uniform standards by placing in the same sub-section and in the same sentence with its power to coin money the power to establish the use of standards of

weight and measure. This covers the subject completely, as every standard for whatever use is some form of measure. Congress has never availed itself of this power, except in a single instance, namely, the establishment of the Troy pound for the purpose of coinage, although it has legalized many suggestions that have been made.

Commercial electrical standards in the main have been unified and in a number of instances legalized. The various electrical associations of America and Europe have by national and international conferences brought about a more or less definite uniformity of standards, although there are still a great many very faulty instruments used for the measurement of these quantities. Unfortunately such instruments are often placed in the hands of those who have no knowledge of their manipulation and thus give unreliable results; yet, on the whole, in electrical fields the instrumental, or measurement side has approximately kept pace with production. In most of the cities there is a regulation of the electrical meters in use. Furthermore the production of electrical energy is in the hands of a limited number who are more or less closely checked unofficially by those who have a technical knowledge or interest in the subject of electricity.

Gas has been the subject of control by the states for a great many years and there is a large degree of uniformity in the standards by which gas is sold. There is, however, no systematic or rigorous inspection of gas meters by competent officials, although this would be highly desirable. Furthermore, the inspection of the gas meters is in many cases of secondary importance to the inspection of the heating and lighting qualities of the gas, which are the two factors which should be primarily considered.

The measurement of water has been and is today an example of the greatest lack of uniformity and lack of accuracy in methods of measurement. Statistics of water supply in cities are largely a matter of guess work except in those instances where reliable meters are used. The majority of cities still sell water by the so-called "flat rate," and the con-

sumption per individual or per square foot of property in any particular section is not known, though such data is of great importance in the consideration of the size of distribution mains alone. The water sold for irrigation purposes is measured largely by the "miner's inch" and the absolute lack of reliable methods of measuring water in the open ditch has been the cause of a great deal of litigation and a great deal of injustice. There is today no satisfactory instrument or set of instruments for measuring this very important and highly expensive commodity.

In this paper I do not wish to take up data relative to other forms of measurement than those of the simplest kind, namely, the fundamental units of length and mass used in the every-day transaction of trade. That there is no intelligent or systematic supervision of the weights and measures which is perhaps the most necessary and simplest form of supervision is rather remarkable. That there is a necessity for intelligent legislation upon the subject by the Government of the United States and coöperation among the individual states can be readily seen by considering some of the legal weights in pounds per bushel of various commodities. Of apples 50 lbs. are required for a bushel in New Jersey and Ohio, and only 45 lbs. in Idaho and Washington; 57 pounds constitute a bushel of onions in Wisconsin, New York and fourteen other states, while 52 pounds constitute a bushel in Massachusetts, 50 pounds in Rhode Island and 48 pounds in Indiana; 60 pounds constitute a bushel of pears in Florida and 45 pounds in Oregon; 60 pounds constitute a bushel of potatoes in New York and 56 pounds in the neighboring state of Pennsylvania. Statistics as to the number of bushels of a certain commodity, even by legal weight in the individual states, are an unfair and unreliable basis of comparison.

The manufacturers of various food products have within recent years acquired the habit of putting up their commodities in a manner which though, perhaps, it may be claimed to be superior from the standpoint of sanitation is far from being conducive to accuracy as to quantity. They have used bottles,

jars, boxes, cartons, pails, all of which vary with the individual manufacturer and are absolutely indefinite; though, unfortunately, they are still used for the purpose of stating the production in terms of quantity, especially, by the dealer who is selling the commodity. As an example, the majority of dealers claim an annual sale of so many pounds of lard, placing the so-called No. 5 pail of the packers as "5 lbs." of lard which introduces an error of about 12 per cent. inasmuch as a No. 5 pail contains only about 4 lbs. and 6 ozs. of lard. The same applies to butter sold in pails and a good deal that is sold in prints, also prepared meats in glass jars or tin cans, crackers in packages, or breakfast foods and cereals in cartons. This should be extended properly to a number of bottle goods, such as a great many syrups, oils and beverages, some so-called "quarts" of olive oil being less than a quart by 46 per cent. Various brands of beer sold in bottles contain only 13 ounces to the pint instead of 16 ounces. Any data, and a good deal of the data are still given in terms of quantity, based on the assumption that these bottles or packages are of a certain definite quantity necessarily give unreliable results. This is not the time nor the place to enter into the further very important question of unfair competition produced by such methods of handling commodities or the question of loss to the public or in how far this enters into the high cost of living.

Aside from the above considerations there is a very interesting phase of the correctness of the apparatus used in the final transactions of trade, namely, in the hands of the retailer selling directly to the consumer. In order to ascertain the prevalence of incorrect weighing and measuring devices, incorrect weights, scales and capacity measures, an investigation was made testing such devices along the main thoroughfare of trade in the City of New York. The following tables give the results of such inspection in the Boroughs of Manhattan and the Bronx.

Business.	Number of places visited.	Scales.								
		Correct.	Faulty.							Otherwise Faulty.
			Fast.			Slow.				
			Loss to Customer.			Loss to Dealer.				
			3%	10%	X	3%	10%	X		
Grocers.	39	24	14	13	7	.....	1	.....	11	
M. markets.....	84	119	17	39	16	.....	4	.....	2	
Fish.....	29	24	4	4	28	3	2	.....	5	
Miscellaneous..	152	42	20	28	39	1	1	.....	38	
Totals.....	304	209	55	84	90	4	8	.....	56	

Correct, 209. Percentage of total 41.30
Incorrect, 297. Percentage of total 58.70
Percentage of total that causes a loss to the customer.....56.32
Percentage of total that causes a loss to the dealer..... 2.37

Business.	Number of places visited.	Weights.								
		Correct.	Faulty.							Otherwise Faulty.
			Under.			Over.				
			Loss to Customer.			Loss to Dealer.				
			3%	10%	X	3%	10%	X		
Grocers.....	39	17	73	14	3	1	.....	.....	.....	
M. markets.....	84	.....	.....	.....	.....	.....	.....	.....	.....	
Fish.....	29	.....	.....	.....	.....	.....	.....	.....	.....	
Miscellaneous..	152	27	5	5	2	6	.....	.....	.....	
Totals.....	304	44	78	19	5	7	.....	.....	.....	

Correct, 44. Percentage of total.....28.75
Incorrect, 109. Percentage of total.....71.25
Percentage of total that causes a loss to the customer.....66.66
Percentage of total that causes a loss to the dealer 4.59

Business.	Number of places visited.	Measures.							
		Correct.	Faulty.						Otherwise Faulty.
			Under.			Over.			
			Loss to Customer.			Loss to Dealer.			
			3%	10%	X	3%	10%	X	
Grocers	39	9	2	3	17	.....	.....	.....	.....
M. markets	84	1	.....	2	8	.....	.....	.....	.....
Fish	29	.....	.....	.....	.....	.....	.....	.....	.....
Miscellaneous	152	18	10	15	36	.....	.....	.....	35
Totals	304	28	12	20	61	.....	.....	.....	35

Correct, 28. Percentage of total.....17.94
Incorrect, 128. Percentage of total.....82.06
Percentage of total that causes loss to the customer.....82.06
Percentage of total that causes loss to the dealer..... 0

The results of the investigation of the conditions in the other Boroughs were similar. Briefly summarizing for the whole of New York City: 617 places of business were visited, 1,227 scales, 1,219 weights, and 511 measures were tested. Thirty-four and nine tenths per cent. of the scales were 10 per cent. or over short, 15.7 per cent. of the weights were 10 per cent. or over short, and 50.1 per cent. of the measures were 10 per cent. or over short. These shortages are of vast economic interest to the purchasing public in their daily consumption of food, and at the same time show the necessity of an efficient supervision so as to reduce the error in order to get reliable statistics as to the amount of such consumption. It must be borne in mind that in not 5 per cent. of the cases is allowance made by the vendor for the inaccuracy of the weighing or measuring device. The above data only take into consideration the actual falsity of apparatus and do not take into consideration the actual apparatus which would still further increase the percentage of inaccuracy.

A similar investigation of the apparatus used by public institutions for checking the quantity of supplies received, upon which checking is based directly the payment of the

same, revealed the fact that in twelve city charitable institutions in New York City about 49.16 per cent. of the scales and 16.96 per cent. of the weights were incorrect and of nine state charitable institutions 52 per cent. of the scales were incorrect and 40 per cent. of the weights. Such inaccurracy must necessarily be to the detriment of the city or state or to the detriment of the one furnishing supplies. Any data as to the amount of supplies consumed would necessarily be incorrect.

Some of the ordinary daily articles of consumption are assumed to be of a certain definite quantity. Butter, particularly when sold in prints is always quoted as "per pound," and statistics on the production of butter by creameries are in the form of so many pounds per annum. In many cases these prints of butter present a shortage of one ounce, or 6¼ per cent., which in itself is a fair percentage of profit. Investigation of forty-four creameries in the State of New York showed that the percentage of the total of all of the prints tested, namely, 435, was 1.26. These represented creameries in different sections of the state and all the weighings of the butter were made while the prints were still dripping wet so that there could be no claim of shrinkage. If this same percentage holds throughout, which is fair to assume, it would make considerable difference in the statements of the total production of butter.

Another article of daily consumption in which the consumer is vitally interested is milk which for assumed sanitary reasons is being more and more largely sold in bottles. In order to test the accuracy of bottles, treating them simply as measures, and testing bottles of twenty-one milk dealers, it was found that 60 per cent. of the quart bottles and 83 per cent. of the pint bottles and 61 per cent. of the half-pint bottles were short, varying from a small amount to a maximum shortage of 2.6 per cent. for quart bottles and 7.8 per cent. for pint bottles.

Bread which is regulated in practically every other country is with the exception of a few isolated instances not regulated in the United States, there being no standards as to what constitutes a loaf of bread. As a result loaves of

bread are of all possible sizes. Taking, of course, only those which are made of wholesome material it was found that of bread baked in forty-eight different bakeries 45.2 per cent. were under 14 ounces which was the size claimed by a number of bakers to be standard size for a loaf of bread; 83.3 per cent. were under 16 ounces, the 16 oz. being claimed to be the standard size of a loaf of bread by the majority of the consumers and the retailers. Another very interesting fact was developed,— that in many of the larger bakeries it is cheaper to purchase two five cent loaves of wheat bread than one ten cent loaf, in short — that the bakers give a larger quantity for ten cents when distributed in two loaves than when put in one loaf.

Without giving further data relatives to the shortages that occur in the sale of coal, wood, thread, dry goods, tobacco, liquors, etc., it seems very necessary that the federal government should exercise the authority given it by the constitution to regulate the weighing and measuring devices used in interstate commerce and that the individual states should exercise their police power in supervising and regulating the measuring and weighing devices there used, as well as to establish standards of quantity and standards of measurement.

The objectors to any form of regulation or supervision of this kind all declare it to be paternalism and interference with the liberties of the people, but it is an incontrovertible fact that the people have a right to such supervision and that the individual cannot and should not have to protect himself by individually checking the standard of measurement, whatever it may be, whether, the electricity consumed, the gas burned or the accuracy of the physician's thermometer or the quantity of coal delivered or the quantity of food supplies purchased.

Some of the figures in the foregoing have been changed so as to incorporate the results of further investigations. It is worthy of note that since writing the above the legislature of the State of New York has passed broad legislation which will result in an efficient and systematic inspection and supervision of the weights and measures used in trade.

ORGANIZED LABOR AND THE RECENT ADVANCE IN PRICES.

By Lewis H. Haney, *Professor of Economics, University of Texas.*

The question whose answer is sought in the following paper concerns the relation between particular prices and particular wage groups. It is perhaps axiomatic that there is no direct causal relation between general wages and general prices; but when one turns to the problem presented by particular prices the matter is less simple. Here anything which limits the supply of a commodity relatively to the demand for it may cause a rise in its price, whether this be monopoly or increased cost of production or absolutely limited supply. Here is the field for increased labor costs and expenses to express themselves. Do employees in the building trades monopolize their crafts? Then, the number remaining the same, the price of houses may be raised. Does it cost more labor to produce a bushel of wheat? Then that bushel tends to bring more on the market. And if the supply of money media is increasing prices will tend to rise at such points first. Probably, when the question is put, what is the cause of the recent high prices, no answer is so frequently in the mouth of the non-scientific man as "organized labor."

In examining this answer it will be well first to analyze the theoretical possibilities and then to bring such statistics as are available to bear.

One way of putting the case is this: If wages are raised relatively to productivity, the price must be raised or the capitalist enterpriser will lose. I say "raised relatively to productivity," for it will be noticed that this may take place in one or both of two ways: either the rate of wages may be raised, productivity remaining the same; or productivity may be decreased, the rate remaining unchanged. Let us consider the latter event first.

Now labor's productivity, equal opportunity being assumed, is reduceable to three factors: energy, time, and skill. If a workman puts forth great effort for a maximum length of time* on a high plane of skill he is very producive. If any one of the three factors is lowered or diminished, his productivity decreases. Of course a similar result would come from poorer opportunities in natural environment, or a lower margin. This would not be attributable to labor, however, and is mentioned only that it may be borne in mind as an alternative or modifying factor.

Has organized labor, then, decreased productivity relatively to wage? If so, through which of the preceding factors does it operate? As to the energy factor, I am unable to get any data which can be taken as proof. It is a matter of common knowledge that the "go-easy," "make-work" or work-fund idea has been widespread and still exists. We know that laborers work harder in years of depression than in boom periods, etc. But these ideas and moral characteristics are not confined to union laborers, and are more and more disclaimed by union leaders. It seems that all that can be said is: (1) that if the go-easy policy obtains among *all* laborers equally prices will not be affected by it; (2) that if in particular trades laborers have adopted this policy within recent years, recent price advances in such trades might be due to it; (3) that we have small reason, however, to lay the exercise of this policy at the door of organized labor, and none for attributing to it the recent rise in prices,— for there is no evidence of any recent change in this regard.

More evidence can be obtained as to the second or time factor. There has been a steady decrease in hours throughout the decade, the index number for 1898 being 99.7† while that for 1907 was 95.0,—a decrease of 4.7 per cent. The demand for and prevalence of the eight-hour day is greater than ever before. To the extent that this means decreased productivity generally, it means less product per laborer, and might mean

*From a long-time point of view the working day would not be so long as to shorten the total working period of the laborer's life unduly.

†Average for 1890–99=100.

a decrease in the "national dividend," or "social income." This would be a misfortune, but it would produce no change in general price levels. Only when confined to particular industries or trades, and not generalized, could exchange values and prices be affected. What, then, are the facts?

The average index number of hours per week for sixteen relatively weakly organized industries was 96.2 in 1907. In the same year the average for eight strongly organized industries was 93,—a difference of 3 per cent. in favor of the stronger unions. When it is observed that such strongly unionized workers as cigar makers, and blacksmiths and horse shoers have index numbers of 99.5 and 94.1 respectively; while common laborers in foundry and machine shops, lumber, and paper and wood pulp have 95.3, 95.2, and 94.5, respectively, the difficulty of drawing conclusions is apparent.*

I conclude that on the whole organized labor has forced greater reductions in hours than have occurred in the case of unorganized laborers and that other things being equal this would tend to raise the particular prices of the goods they produce. Whether other things have been equal and prices of organized labor's products have risen at the expense of other products will appear in a moment.

As regards the third factor, skill, I must be content with

*Table showing relative hours for weak and strong organizations.

WEAK.		STRONG.	
Agricultural Implements,	96.3	Bread,	91.6
Boots and Shoes,	96.0	Blacksmith and Horse Shoeing,	94.1
Carpets,	98.6	Building,	90.6
Cars,	95.9	Harness,	95.9
Cotton Goods,	96.8	Brewery,	87.0
Electrical Apparatus and Supplies,	93.3	Stone and Marble,	93.6
Foundry and Machine Shop,	94.6	Printing,	92.4
Gas,	96.1	Cigars,	99.5
Glass,	98.6		93.0
Hosiery and Knit Goods,	97.7		
Iron and Steel { Bessemer	90.5		
Iron and Steel { Blast	100.6		
Leather,	99.9		
Paper,	89.9		
Silk,	97.6		
Woolen Goods,	98.0		
	96.2		

observing that to the extent trade unions force employers to use skilled men where unskilled men ["handymen," "helpers," etc.] could do the work, they may virtually decrease productivity relatively to wage,—or, more simply, increase or prevent a decrease in cost. To what extent this is done there is no means of learning. (Of course, if skill be considered alone or absolutely and not in relation to cost it will generally be admitted that on the average the most skilled men of the craft will belong to the union and unions rather foster than decrease skill.)

As to the movement of the other factor in "relative productivity," i. e., rates of wages, some inferences may be drawn from the statistics of wages and hours of labor from 1890–1907.* From the table on page 73 I conclude that of those rises in wages which are above the average rise there are more in case of weakly organized than in case of strongly organized trades. Furthermore, the average of relative wages in 1907, though it seems somewhat lower for weakly organized industries than for strong, shows no such difference as to indicate any effect of unions on particular prices, the average index number for sixteen weakly organized trades was 125; for eight relatively strong trades it was 130.†

*Bulletin, U. S. Bureau of Labor, No. 77.

†Table showing relative wage rates for weak and strong unions.

WEAK.		STRONG.	
Agricultural Implements,	130.9	Bread,	128.9
Boots and Shoes,	124.3	Blacksmith and Horse Shoeing,	126.4
Carpets,	117.1	Building,	144.6
Cars,	124.4	Harness,	123.5
Cotton Goods,	157.5	Brewery,	132.9
Electrical Apparatus and Supplies,	122.6	Stone and Marble,	125.7
Foundry and Machine Shop,	121.4	Printing,	126.8
Gas,	129.4	Cigars,	132.4
Glass,	123.5		
Hosiery and Knit Goods,	133.4		130.1
Iron and Steel { Bessemer	119.8		
Iron and Steel { Blast	132.6		
Leather,	111.8		
Silk Goods,	116.9		
Slaughter and Meat Packing,	116.0		
Woolen,	131.9		
	125.8		

Can we not infer, then, that if other things were equal the tendency of labor unionism would be, through affecting relative productivity, to cause higher prices in the organized part of the industrial field than in the unorganized, and also in the strongly organized as compared with the weakly organized part; and that the scanty statistics as to hours and wage rates show some evidence of this tendency, though no such marked differences as might be supposed appear.

Inasmuch as positive conclusions from this examination of the possibilities and actualities of the more immediate means of affecting prices are so meagre we must seek results elsewhere. We have concluded that the only way in which laborers can effect a rise in prices is to decrease relative productivity in a particular industry or group of industries through one or more of the factors just briefly discussed. But the ability to make such a decrease depends upon the strength of organization of the union; just as, in so far as prices are concerned, a union can work only through a change in relative productivity, so the capacity to affect relative productivity and prices depends upon the strength of the union.

To be sure, we cannot measure changes in relative productivity. Merely a few deductions have been suggested. Therefore, when a presence or absence of such price changes as altered relative productivity would effect is found corresponding with the strength of labor organization, we cannot be sure as to a causal connection. But knowing something of the possibility and tendency of labor organization in this regard some deductions may be drawn when union strength is found coupled with relatively low prices, and *vice versa*. In a word, conclusions negative so far as causes are concerned may be drawn,— and for the problem concerning the relation of organized labor to prices negative conclusions are all that is necessary.

Take, then, as the most general comparison available that between raw materials and manufactured goods. It matters not that the line cannot be drawn exactly. It is reasonably

certain that trade unions have their great strength in the more skilled manufacturing industries; and, that, with the exception of a considerable part of the bituminous coal producing industry, the laborers engaged in producing raw materials are unorganized or but weakly organized. Yet the latest statistics show advances in the index number of wholesale prices to 125.5 in the case of raw materials and to but 122.2 for manufacturers, the average for 1890–99 being 100.*

More particularly. Farm products head the list of certain great groups at 133.1, the average for all products being 122.8. Next come lumber and building materials at 133. Foods are estimated at 120.6. On the other hand clothes and clothing have risen to 116.9; house furnishings to 114; metals and implements to 125.4; and fuel and light to 130.8,— the last containing many non-union and weakly organized occupations.†

But let us make the comparison still more particular and more direct. Take a group of industries in which unionism is strong and compare those with other industries in which it is weak. Here the difficulty which at once confronts one is that one does not know with any accuracy what the weak and what the strong ones are. And in other cases some knowledge

*Bulletin, U. S. Bureau of Labor, No. 81, p. 209. Wholesale prices, 1890–1908. Relative prices in industries virtually unaffected (directly) by labor organization:

	1908	1907
Cotton	134.8	153.0
Grain	163.0	148.3
Livestock	122.3	129.7
Average farm products	133.1	137.1
Fish	124.9	128.3
Milk	131.4	129.0
Eggs	142.0	141.2
Average food products	120.6	117.8
Jute: Raw	140.4	184.4
Average Misc.	119.9	127.1

†The difficulty here lies in the fact that each of these groups — except the first — contains some strongly organized trades. Conclusions are difficult except as we roughly estimate the relative extent of such organization.

on this point may avail one nothing in that it would be impossible to estimate what part of the product whose price has risen is assignable to the known union. It goes without saying that mere absolute numbers, notoriety, etc., are scant evidence of strength. It is necessary to turn aside for a moment to pick our strong and weak unions.

In the Report of the Commissioner of Labor for 1906 there are sundry tables concerning the number, extent, and success of strikes by states and by industries for the years 1881–1905. On page 33, for example, a table shows the percentage of strikes ordered by labor organizations by industries. I find that in a considerable number of industries a large percentage of strikes is not ordered by such organizations.* This proves nothing for my purposes, it is true. It does not indicate much as to the strength of any particular trade organization in some cases, and is not conclusive in any case. It merely affords some presumption as to the average control of trade unions in a whole industry for a series of twenty-five years,—control over the declaration of strikes.

The result is virtually identical if the percentage of establishments affected by strikes declared by organized labor be taken rather than the number of strikes.

But other tests remain. What proportion of the employes went out on strike when ordered by labor organizations, i. e., how inclusive were the unions? I find that the list of low percentages is almost identical with the list judged to be weak on the preceding basis. Perhaps eight of the industries indi-

*Agriculture.
Brass and brass goods.
Brick and tile.
Canning and preserving.
Car building.
Carpets.
Coal and coke.
Cotton goods.
Cutlery, etc.
Freight handling and teaming.
Gas.
Glass.
Hardware.
Hosiery and knit goods.
Iron and steel.
Leather.
Lime and cement.
Lumber and timber products.
Mining, ore.
Paper and paper goods.
Pottery.
Railroad, canal and road building.
Railroad transportation.
Rubber goods.
Silk goods.
Smelting and refining.
Tobacco: chewing and smoking.
Woolen goods.

cated as weakly organized by the preceding test show greater strength;* while four industries would be added to the weak list.†

I turn to another table which shows the results of the strikes ordered by labor organizations during the same twenty-five years. On the whole there is a very striking similarity in the indications as to strength and weakness. The result of applying the new test is that a question mark must be placed after five out of the list of thirty-two suspected of weakness on the preceding bases;‡ while one additional industry shows weak labor organization in a large percentage of strike failures (street railways).

Finally, there is a table of strikes settled by joint agreement, that is, between officials representing organizations on both sides. Surely those industries in which strikes are largely settled in this way are strongly organized, though perhaps not *vice versa*. Not one of the industries showing any considerable percentage of strikes so settled appears in our list of weak unions, unless it be decided that freight handling and teaming, and coal and coke are weak.§

*Agriculture (69%).
Brick and tile (72%).
Coal and coke (76%).
Frt. handl. and team (63%).
Lumber and timber (55%).
Mining, ore (57%).
Silk goods (51%).
Smelting and refining. (53%) — the last two are obviously not strong.

†Boots and shoes.
Foundry and machine shop.
Slaughtering and meat packing.
Stoves and furnaces.

‡Hosiery and knit goods.
Lumber and timber products.
Paper goods.
Pottery.
Tobacco: chewing and smoking.

However, 91 per cent. of strikes in hosiery, etc., were "unorganized" and 61 per cent. of all strikes failed. Taking all strikes, 49 per cent. failed in lumber, which is above the average; and paper goods and pottery are only average in success on the same basis.

§The most notable, taking number and per cent. into consideration:
Blacksmithing and horseshoeing.
Brewing.
Building trades.
Coal and coke.
Stone quarrying and cutting.

Others are: Men's clothing, freight handling, planing mill products, and cigars and cigarettes.

It may be added that an examination of the statistics published by the Connecticut bureau showing movement of number of employees, wages, and hours by industries may be made the basis for conclusions which support the preceding.

The net result of the inquiry is the inference that the following industries taken as a whole show that labor organization is least effective in them:

Rubber goods.	Iron and steel.
Canning and preserving.	Car building.
Mining, ore.	Carpets.
Cotton goods.	Woolen goods.
Hosiery and knit goods.	Silk goods.
Hardware.	Leather.

Those industries in which unions seem most effective are:

Blacksmithing and horseshoeing.	Building trades.
Men's clothing (and women's).	Gloves and mittens.
Bakery.	Stone quarrying and cutting.
Brewing.	Tin and sheet metal goods.

Coal and coke shows strong in extent and inclusiveness of organization, but rather weak in failures of organized strikes (47 per cent.). It seems clear that labor organization is strong in a considerable part of the bituminous coal area, but weak elsewhere.

Returning to the comparison, to make which we turned aside to inquire into the relative strength of labor organizations in various industries, the conclusion must be drawn that no appreciable effect of organized labor on the rise of particular prices during the last decade is apparent. Comparisons are available for the following weakly organized industries:

	1908	1907
Rope, twine and bagging	121.0	138.0
Cotton goods	115.2	138.0
Carpets	118.9	123.2
Iron and steel	113.9	135.3
Woolen goods	122.7	128.0
Tools	113.6	115.7
Boots and shoes	121.3	125.9
Average relative price	118.0	128.4

The average relative prices for the group of seven is 118 in 1908, and 128.4 in 1907.

I experienced more difficulty in finding strongly organized industries for which prices were available. Among the strongest only two such were found (bread and coal) neither of which affords satisfactory conclusions. But adding five of those which were much stronger than the ones included above, I find that the average relative price for the years 1908 and 1907, respectively, is 119.0 and 119.9.

	1908	1907
Bread	113.6	111.4
Coal	132.7	134.2
Lime and Cement	107.4	104.3
Tin and Sheet Metal	110.5	115.7
Tobacco: Chewing and Smoking	118.3	118.3
Paper	86.7	87.4
Lumber and Timber	164.0	168.6
Average	119.0	119.9

Thus in 1908 the index number of prices in seven weakly organized industries was one point lower than in seven relatively strongly organized ones; while in 1907 it was actually 8.5 points higher. The year 1908 was one of generally lower wholesale prices than in 1907, and it is not without significance to note that the more strongly organized industries showed the smaller decrease.

Massachusetts statistics show that for the year ending September 30, 1907, 98 per cent. of advances in wages occurred without strike. And of these at least 67 per cent. were not the result of any immediate union activity. The amount of the advance, however, was somewhat less than in the 33 per cent. in which unions figured.*

Thus, while pointing out the theoretical possibilities of labor organization in regard to particular prices, the conclusion is to be drawn that there has been no great or widespread effect. Organized labor has reduced hours; but it is not clear to what extent this means decreased productivity, nor has the reduction been confined to organized labor. There seems to have been no great difference in price movements as between weakly organized and relatively strongly organized industries, while the greatest advances have come in industries which are practically unorganized.

*First annual report on changes in rates of wages.

A FORMULA FOR DRAWING TWO CORRELATED CURVES SO AS TO MAKE THE RESEMBLANCE AS CLOSE AS POSSIBLE.

By Francis Todd H'Doubler.

One of the problems with which statisticians are confronted almost daily is that of drawing two curves in such a way that the correspondence, so far as the eye can judge, shall be as close as possible. The statistician, as a rule, simply experiments with his scales, varying the two units until the curves fall into some sort of correspondence, abandoning his graphical illustration if the appearance of the two curves is not such as to warrant a positive conclusion. As is shown below, it is possible to derive a formula which will, without imposing too much labor upon the statistician, make this resemblance as close as could be desired. If, for instance, we are comparing yearly marriage rates with yearly prices of wheat, the two groups will be brought into closest correspondence if, after plotting the marriage rate, we multiply each annual price of wheat by the expression $\frac{\Sigma x_i\, y_i}{\Sigma y_i{}^2}$, where $x_1, x_2 \ldots x_n$ represent the annual marriage rates, and $y_1, y_2 \ldots y_n$ represent the annual prices of wheat. By closest correspondence, in the above statement, is meant such juxtaposition of the two curves as will make the sum of the squares of the distance between the two curves at the successive years a minimum. This definition of closest correspondence has much to be said in its favor on general mathematical grounds. If, however, we define closest correspondence as such juxtaposition as will make the sum of the distances between the two curves regardless of sign, a minimum, the problem becomes much more complex and the solution very laborious. In the great majority of instances, moreover, the adjustment of the two curves by the two methods would probably be practically identical. The problem has possibly been worked out before, but has never come to my attention; and the utility of the formula in practical statistical work warrants its reproduction, even if it has been previously published.

Let $x_1, x_2, x_3 \ldots x_n$

and $y_1, y_2, y_3 \ldots y_n$

be the statistics of two correlated phenomena, where x_i and y_i correspond to the same value of the variable. Thus x_1 and y_1 might be respectively the marriage rates and the average wheat prices of a certain year; x_2 and y_2 the rates of the next year, and so on.

If these figures are plotted, two curves are obtained which show in a way the correlation existing between the two phenomena. We wish to make the two diagrams fit as closely as possible.

Plot the x's just as they occur. Then multiply each y by the quantity e, and plot the resulting figures. We have the two sets

$$x_1, x_2, x_3 \ldots x_N$$
$$ey_1, ey_2, ey_3 \ldots ey_N$$

Consider the sum of the squares of the differences between corresponding x's and y's. This quantity is

$$(x_1-ey_1)^2+(x_2-ey_2)^2+\ldots+(x_N-ey_N)^2$$

or, in short

$$\sum_1^N i(x_i - ey_i)^2$$

It must be remembered that x_i and y_i are fixed numbers. Evidently

$$\sum_1^N i(x_i - ey_i)^2 = f(e)$$

is a function of e—, that is, it varies with e. Let us seek that value of e which causes this function to take its minimum value.

First of all we note that $\Sigma(x_i - ey_i)^2$ is the sum of squares and hence of positive quantities. Consequently it is always positive. Moreover, it increases beyond limit as e increases. Therefore, when the derivative of this function is equated to zero we have the condition for a minimum (by differential calculus).

Differentiating and equating to zero, we have

$$-2(x_1-ey_1)y_1-2(x_2-ey_2)y_2-\ldots-2(x_N-ey_N)=0$$

and solving for e

$$e=\frac{x_1y_1+x_2y_2+x_3y_3+\ldots+x_Ny_N}{y_1^2+y_2^2+\ldots+y_N^2}$$

or

$$e=\frac{\sum_1^N i x_i\, y_i}{\sum_1^N i y_i^{\,2}} \qquad (1)$$

When the y's are multiplied by this value of e the sum of the squares of the differences $(x_i - ey_i)$ is a minimum, and the x— curve and the modified y— curve are sure to fit closely.

In case the y's are unaltered while the x's are all multiplied by e^1, the value of e^1 giving a minimum to $\Sigma(y_i - e^1x_i)^2$ is similarly found to be

$$e^1=\frac{\sum_1^N i(x_i\, y_i)}{\sum_1^N i x_i^{\,2}} \qquad (2)$$

If $x_i = y_i$, then both e and e^1 are equal to unity.

One might ask why it would not be better to consider merely the differences ($x_i - ey_i$) regardless of sign, and to seek the value of e causing the sum of these differences to assume its minimum. In other words, consider the sum of the absolute values,* $|x_i - ey_i|$, — *i.e.*

$$|x_1 - ey_1| + |x_2 - ey_2| + |x_3 - ey_3| + \ldots + |x_N - ey_N|$$

or

$$\sum_1^N i \; |x_i - ey_i|$$

This function is always positive. Its graph is a broken line with N corners—a corner being located above each of the abscissæ, $e = x_i / y_i$ (i=1 . . . N). One of these N ratios gives

$\sum_1^N x_i - ey_i$, its minimum. It must be found by actually substituting each ratio and seeing which gives the above sum the smallest value. There is no general formula applicable here, since the derivatives at these corners are all discontinuous. This method involves so much tedious work that the first method, with its general formula, is much more practicable, as well as preferable.

*In mathematical notation $|x|$ is used to indicate the absolute or numerical value of x. Thus $|-10| = |10|$.

REVIEWS AND NOTES.

THE AMERICAN FRATERNAL EXPERIENCE COMMISSION.

At the annual convention of the Associated Fraternities of America, an advisory federation of about fifty fraternal benefit societies operating in the United States and Canada, held in Chicago in August, 1908, the following resolutions were unanimously adopted, to wit:

"*Resolved*, That an American Fraternal Experience Table of Mortality, constructed from data now available or easily accessible, by actuaries of recognized professional standing is one of the most urgent needs of the hour in the solution of the rate problems confronting so many fraternal beneficiary societies.

"*Resolved*, That a commission of three be named by this convention to devise ways and means and to coöperate with like commissions from similar organizations or individual societies, for the construction of such a table."

The Commission as thus authorized was duly organized as follows:

President, LEE W. SQUIER, Philadelphia, Pa.
Vice-President, J. L. ROSE, Waterloo, Iowa.
Secretary-Treasurer, C. H. ROBINSON, Chicago, Ill.

An actuarial staff, consisting of five members, was chosen by the Commission, as follows:

Actuary-in-Charge, MILES M. DAWSON, F. I. A., New York City.
Consultant, WILLIAM FRANCIS BARNARD, F. A. I. A., Syracuse, N. Y.
Consultant, ABB LANDIS, Nashville, Tenn.
Consultant, GEORGE DYBE ELDRIDGE, New York City.
Consultant, SIDNEY H. PIPE, F. A. S., Toronto, Can.

Letters were then written to the various fraternal societies in the United States and Canada, outlining in brief the scope, methods and purposes of this Commission. From the replies received it was found that the amount of data available was so vast and extensive that not only could a table of general mortality be constructed; but also many other tables which would be invaluable not only to the fraternal insurance world but to the science of life insurance generally and to the general philanthropic movement for the improvement of the economic conditions of wage-earners.

These tables, as finally projected by the Commission, are as follows:

I. General Mortality Experience.
II. Mortality Among Male Lives.
III. Mortality Among Female Lives.
IV. Mortality Experience As to Residence.
V. Mortality Experience by Occupations.
VI. Experience As to Lapsation.
VII. Temporary Disability Experience.
VIII. Partial and Permanent Disability Experience.
IX. Total and Permanent Disability Experience.
X. Tabulated Causes of Death.
XI. Tabulated Causes of Disability.

Accompanying the tables showing mortality and disability experience there will be commutation columns, monetary tables and other mathematic and scientific deductions from this wide investigation that will make the Commission's published reports veritable storehouses of information on sociologic and economic problems.

Each society entering into this movement fills out an information card for every person who has ever become a member of it. These cards are then sent to the office of the Actuary-in-Charge, there to be sifted, sorted and counted for the various tables for which they are pertinent. The society then gets its cards back, together with sheets showing its ungraduated experience along the various lines of the investigation; which, at slight actuarial expense, can be thrown into graduated form and thus make the foundation of the society's permanent experience tables.

The cards now completed or in process of writing on the form designed by the Commission number about two millions; and about three millions more are offered by societies on their own forms, which can easily be adapted to the Commission's use, thus placing some five million of lives under observation in the construction of the eleven tables projected.

When it is remembered that the Actuaries', or Combined Experience Tables, published in 1843, were based upon only 62,537 lives; and that the American Experience Table, published in 1867, was merely an adaptation of the Actuaries' Table to the experience of the Mutual Life Insurance Company of New York; and that these two tables combined — both of which have been generally accepted as standards — were made up of less than one-fiftieth of the number of lives that will form the basis of the general mortality table projected by this Commission; and when it is remembered that the English Healthy Districts Males Tables were based by Dr. Farr on the census returns of 1851 and deaths during the five years from 1849 to 1853, in

sixty-three districts of England and Wales, to the number of 87,345 in a total population of 996,773, and were not made up of insurance exposures at all; and when it is remembered that the other ten tables outlined in this movement, besides the general mortality table, are, many of them, along lines never heretofore followed with scientific accuracy,— then it will be seen what a tremendously important and far-reaching investigation this Commission has planned in the interests not only of safe and sound protection of life and health, but also in the interests of economic and public welfare.

The counting of the cards has begun in the office of the Actuary-in-Charge at New York and from now on the tabulations will proceed as rapidly as the great volume of the data will permit.

The Commission is now extending its operations so as to include the experience of the numerous Trades Unions connected with the American Federation of Labor. In a recent letter to the President of the Commission, President Samuel Gompers of the American Federation says:

"I agree with you entirely that to ascertain accurate data of the disability and mortality will prove of great advantage in securing intelligent, beneficent legislation. To determine the vocational disabilities and deaths will be the means of endeavoring to correct evil causes, rendering injury and disability less frequent and will help the prolongation of the lives of our workers."

Already many of the federations have signified their hearty approval of this movement and indicated their willingness to furnish the data asked for by the Commission.

Every safeguard known to actuarial and statistical science will be enforced by the Commission to place the methods of this investigation above possible criticism and to make the results authoritative and permanent.

LEE W. SQUIER.

The Standard of Living Among Workingmen's Families in New York City. By Robert Coit Chapin, Ph.D., Professor of Economics in Beloit College. Russell Sage Foundation Publication, 1909. (New York: Charities Publication Committee, 105 East Twenty-Second Street, pp. xv., 372, $2.00, postpaid.)

This study was carried out by Professor Chapin as secretary of a special committee on standard of living, appointed by the New York State Conference of Charities and Correction. The main part of the volume is the New York investigation but there are also included a valuable introductory chapter on Workingmen's Budgets in Statistical Literature, a translation of one of the individual studies (1857) that appears in Le Play's *Les Ouviers Européens*, a partial bibliog-

raphy, the report of the Committee on Standard of Living, a report on nutrition by Prof. F. P. Underhill, the committee's expert, and reports on the standard of living in nine other cities and towns in New York State; these last reports are, however, based upon very meagre data.

The data for the investigation in New York City were gathered mainly by paid agents during the summer of 1907. Detailed schedules were secured from over 600 families but only 391 schedules for families of four, five or six members, with incomes ranging from $400 to $2,300 per annum, were used. The main analysis was restricted to 318 families having incomes between $600 and $1,100. The families from whom information was obtained were, apparently, chosen at random. The schedule is very complete, containing questions as to occupation, mental and physical condition, housing, varieties of food in detail, fuel and light, insurance and savings, clothing, health, taxes and contributions, recreation and education.

The 318 families were classified by nationality as follows: 67 United States, 39 Teutonic nations, 24 Irish, 28 colored, 14 Bohemian, 57 Russian, 32 Austro-Hungarian, etc., and 57 Italian. The occupations were mainly those of the less skilled employments, such as laborers, teamsters and garment workers. The tables throughout give the data concerning the 318 families according to nationality and by income classes of $100 each. Averages and percentages for the various nationality and income groups are presented. The classification of families by incomes is necessary and illuminating, but it does not seem that the persistent classification by nationality is of much, if any, significance.

There are eight nationalities and five income classes, making forty divisions in all, for 318 families. Consequently twenty-eight divisions contain less than ten families each and eleven divisions contain less than five families each. These numbers are too small to base any reliable conclusions upon concerning the influence of nationality. In order to avoid this criticism Professor Chapin has combined the nationalities into two groups, which expedient is not a satisfactory method of meeting the difficulty. We are told little of the extent of variation *within* the nationality groups so that we cannot judge as to the significance of the differences between averages for different nationalities.

The earnings of the father of the family constitute the main source of family income, although there were additions from the earnings of mother and children, from lodgers, and from other sources. The analysis of earnings would have been more illuminating had families been classified according to the source of income and had the percentage of each variety of income (mother, children, lodgers, etc.) to the total income been determined for *those families receiving such additions*. The average income and the percentage of total in-

come received from lodgers does not indicate much where families having no lodgers are combined with families having lodgers. (See diagram 5). Then, too, would not those families keeping lodgers be apt to spend a greater percentage of income for rent and food?

There appears to be some inconsistency in Table 81, which gives the average expenditure for clothing for each member of the family by income groups, with and without gifts of clothing. In the $600–$699 group the average expenditure for each girl in the forty-seven families not receiving gifts is $10.48; the average in the twenty-five families receiving gifts is $8.01, while the average in the combined seventy-two families is *more than in either case, i.e.*, $10.90. The average expenditures per boy or per girl per family are not weighted by the number of boys or girls in each family to get the average for all families in each income group.

From the preceding criticisms I would not have the reader infer that the investigation and the statistical presentation are not of value. The investigation is one of importance and the statistical analysis throws much light upon significant questions. Dr. Chapin has defined standards of over-crowding (more than 1½ persons per room), of under-feeding (an expenditure of less than twenty-two cents per man per day), and of under-clothing (an expenditure of less than $100 per family of five per year). Of the 318 families having incomes between $600 and $1,100, fifty per cent. are over-crowded, forty per cent. are under-clothed, and twenty-three per cent. are under-fed. An examination of the statistics of the under-fed, under-clad, and over-crowded families as to whether their expenditures more or less usually come within their incomes than do the expenditures of other families leads Dr. Chapin to the significant conclusions that "the families that make both ends meet on less than $600 are living far below the normal standard" and that "a comfortable margin of income over expenditure is first possible with an income between $800 and $900" (p. 230). Of the 318 families only 36.5 per cent. have a surplus. These facts suggest the question, is it not economically undesirable, is it not improvident, for those families to save where "an even balance or a surplus can be attained only by curtailing expenditures for necessaries below the point of meeting the requirements of healthy existence" (p. 232)?

The prosecution of similar investigations in various parts of the country would enable economists and sociologists to come to most certain conclusions regarding such questions as the possible source of savings, the need for laws to standardize grades and packages of food and the advisability of old age pensions and insurance against accident and sickness.

WARREN M. PERSONS.

Dartmouth College.

Women and the Trades. By Elizabeth Beardsley Butler. Russell Sage Foundation Publication, 1909. (New York: Charities Publication Committee, 105 East Twenty-Second Street, pp. 440, $2.00, postpaid.)

This volume is the first of a series of six giving the findings of the Pittsburgh Survey of 1907–'08. It is a study of those industries and trades employing women, with reference to the wages, hours of work, and general conditions of labor of the workingwomen employed. Each of 448 shops and factories were visited and the data were secured from employers, employees, and other persons familiar with the trade. In several cases access to the payrolls was secured. The conditions in each trade and industry are discussed separately, and a general resumé is given in the closing chapters of the report, followed by a bibliography of works on women in industry.

The data are presented dispassionately. The multitude of facts concerning the conditions of employment of women are given in such an effective manner that the reader can draw definite conclusions as to what laws need more thorough enforcement and what additional legislation is required. A summary of the rates of wages received by 22,185 women in twenty-seven different classes of work (p. 338) shows that 62 per cent. receive less than $7.00 per week, 21 per cent. from $7.00 to $7.99, and 17 per cent. receive $8.00 or over. Of the 7,540 women in mercantile employment 73 per cent. receive less than $7.00 per week. Miss Butler says that not one of the workingwomen or others interviewed "was willing to consider $6.00 a living wage. They agreed that the minimum below which a working girl cannot live decently and be self-supporting in Pittsburgh is $7.00 a week" (p. 346). When we consider that many of the occupations are seasonal, and that nearly all have their dull seasons lasting several weeks, thus causing irregularity in employment, we can appreciate something of the economic pressure felt by these working girls.

Payment by the piece is the rule. Numerous instances are given where the piece rate was cut after the girls had reached their maximum speed and had secured earnings approximating $1.50 per day (pp. 218 and 263). There are no unions to establish a "*ca canny*" policy and make it effective. Much evidence of over-speeding is given. A manager of a large stogy factory holds that in his industry "No girl can keep up her pace more than six years" (p. 96). Miss Butler concludes that the stogy industry "is taking young, undeveloped girls, lifting their speed to its highest pitch, and wearing them out" (p. 96). In telephone work "the life of an operator" was given by the managers to be from eighteen to twenty months in one company and fifteen months in the other (p. 291).

A ten-hour working day is the rule. Over-time and night work are common in many industries, such as laundry work, canning, and

confectionery. The conditions surrounding workers in laundries, bakeries, the garment trades, laquering of metal, etc., were found to be unhygienic and unsanitary in many instances. The favorable conditions in certain factories showed that work need not be carried on in a manner unhealthful to the employees. In many factories the machinery is not properly guarded and the inevitable accidents happen (p. 234).

"*Women and the Trades*" is a scientific and effective presentation of facts based upon a careful study of conditions. Such a study is a necessary preliminary to an intelligent reform movement.

WARREN M. PERSONS.

Dartmouth College.

Housing Reform, A Handbook for Practical Use in Modern Cities. By Lawrence Veiller. Copyright, Russell Sage Foundation, New York, 1910.

There was a time when to appear scientific in matters sociological it was not uncommon to apply long and involved terms and complicated systems of classification to the concerns of every-day life in order that the mark of "investigation" might be upon them. The scientific purpose of such study is now clearly recognized to be to place before the average citizen the facts about himself and his community in such guise that they interest him to the point of action and provide him with the kind of information which assists him to intelligent activity in the progressive welfare of his town or city and thus of his country. Writers on social questions, therefore, present their material not only to those like themselves technically equipped but to the general public. Material must be reduced almost to the point of classical simplicity.

In "Housing Reform" Mr. Veiller has this essential purpose in presenting his information on city housing clearly before him and has chosen to condense his extensive expert knowledge on the administration of city tenements into a handbook of two hundred pages. Mr. Veiller's book is not merely valuable in its simple and concise statements as to the steps looking to Housing Reform. It will serve almost equally well as a handbook in strategic practical methods for public spirited citizens who seek by private organized action the varied interests of the community.

Mr. Veiller deals only slightly with figures and yet in method of utilizing statistics advantageously so that their value to the general reader is fully available this book is an illuminating statistical document. This is to be found in the closing paragraphs of the chapter on "Model Tenements and Their Limitations" where the argument is clinched by utilizing a few well selected and contrasting figures.

The essential thesis presented by Mr. Veiller is the practical task resting upon society of establishing ways of living for all, in keeping with our common standards of civilization. Nothing so much determines a due degree of decency as adequate housing facilities. "It is eminently the right of the state and of each community to see to it that people are not permitted to live under conditions which make for poverty, sickness, disease and death; which produce pauper, criminal and vicious citizens, whom the state is subsequently called upon to support and reform," says Mr. Veiller, and no one can deal with city conditions for long or become by any means acquainted with communities of working people without reaching such a conviction. One views the hardships placed upon women who keep house in the many varieties of ill-equipped tenements as one of the greatest injustices imposed by one human being upon another. It is only too evident that some method of adequate regulation must be devised since the difficulties and expense of securing proper facilities is more than the powers of the average individual family can manage. What this regulation shall be, under what power of the state it shall be exercised and how it is to be secured, are the questions which Mr. Veiller answers in his handbook. It is his desire to make the mammoth experiment in tenement housing which New York City presents of practical value to the rest of the country, and it is out of his own active participation in the New York situation both as a private citizen in securing the present law and as public official in administering it that the author writes. He produces a program easily adapted to cities of varying sizes and brief advice in a chapter of "dont's" which if followed in the early stages of town growth will reduce the solution of a future tenement house problem to a great degree. Foremost among these injunctions Mr. Veiller indicates his own very human point of view and, as throughout the whole book, his practical wisdom: "Don't let your city become a city of tenements, keep it a city of homes." Stress is laid on the fact that as every city community anticipates growth so each has its present duty to forestall evils if they do not already exist.

This handbook suggests to the uninitiated things to look for, the way to look, and what to do about them. The program calls for: (1) A carefully planned investigation to be made as a basis for action. (2) State legislation, after conditions are known, in keeping with whatever situation the investigation sets forth. (3) Development of public opinion to secure both the law and its intelligent enforcement. (4) Securing effective enforcement by using methods of cooperation with public officials. The field for private effort should be found in dealing with old buildings and their possible improvement rather than in the construction of model tenements.

A person inexperienced in getting things done in a municipality might smile somewhat scornfully at the detailed methods which

Mr. Veiller is at pains to set forth at each step indicated; but a very short period of active effort would quickly demonstrate that it is by such intensive arts in the association with men that things do get done. The chapter on "How to Secure Legislative Reforms" is of particular value in this respect. Mr. Veiller wants housing reform successful and to that end he not only tells what to do but, equally important, how to do it. His suggestions indicate the qualifications necessary for success. They might be summed up as: honest human sympathies; sagacity in asking for public action in the way most likely to secure it; extreme desire to get substantial results; patience in dealing with men and things.

Mr. Veiller will undoubtedly encounter the most adverse criticism in failing to advise the building of model tenements, partly because of a real difference of opinion and, partly also, because of the opportunity which such enterprises suggest of exercising a gentle tyranny toward cleanliness over the poor, a sentimental appeal to those who wish to be kind without assuming the charitable rôle. The experience in New York is, however, convincing, and where model tenements are under consideration a careful study of local conditions in comparison with those of New York should certainly be made. It is possible that in spite of the excellence of Mr. Veiller's judgment circumstances in other communities might make a plan for model dwellings effective. So huge an aggregation of tenements as exists on almost every street in New York very largely precludes the existence of any neighborhood life. Where, however, such local ties do thrive as in smaller cities the moral standards of a whole section may be affected by the activity of a comparatively small group. Similarly the improvement in environment which comes from well built, well equipped tenements would affect tenants and neighbors alike. It is conceivable that a model housing undertaking might thus be made, in conjunction with specially designed social leadership, an effective instrument in establishing standards for normal neighborhood conditions, in which, because of his local identity, the individual thrives to advantage.

However it may be done, the purpose of a wise Housing Reform movement should be to glorify the calling of the landlord. To that end the municipality should exercise a firm hand in sanitary and building regulation. Mr. Veiller's Handbook cannot fail of being of great service in this direction. His careful guidance is recommended to the citizens who know their city conditions to be unsatisfactory or who do not know what the conditions are. Other workers in the field of social reform will find by practical use of his handbook an effective way of expressing their appreciation of Mr. Veiller's service in preparing it and of the Russell Sage Foundation in its publication.

ELEANOR H. WOODS.

An Analysis of the United States Steel Corporation With Chart of Quarterly Earnings Since Its Organization. By Charles Wilson Taintor. Boston, 1909.

This publication by Mr. Taintor consists simply of a folder containing two large sheets printed on one side. The first page contains the accompanying statistical chart, and the second gives a brief analysis and explanation of the curves and figures on the chart.

Mr. Taintor's work is so admirably done and has received such high commendation from the best financial authorities that it seems to the writer well worth while to reproduce his chart and to include a brief review of his analysis for the benefit of those who do not have access to private publications of this sort which have only a limited circulation.

The author explains the construction of his chart as follows: "The quarterly earnings since organization are charted in millions of dollars, the total for each quarter being plotted vertically from the base line, and the points thus established being connected by a black line (this is the highest line on the Chart and indicates the variations in the Corporation's quarterly earnings). Each quarterly total is sub-divided vertically as indicated by colored figures into the various amounts corresponding to the quarterly statements of the Corporation." The chart contains also a broad vertical line divided into the same colors to show the average amounts appropriated to each purpose during the period covered.

It will be noted that Mr. Taintor has used these familiar statistical devices with excellent effect. The scale adopted and the arrangement of colors render it easy to see at a glance not only the total earnings for each quarter but just what was done with them.

It may be worth while to note in passing that the curve indicating the total earnings per quarter exemplifies by its quick and extreme fluctuations the truth of Andrew Carnegie's assertion that the steel industry is "either a prince or a pauper." The great steel corporation has been able to hold prices fairly steady, but it could not or at least did not prevent its own earnings from falling off nearly two thirds in nine months, June, 1903, to March, 1904, or from $36,600,000 to $13,200,000. From this low point the quarterly earnings rose by leaps and bounds to $45,500,000 in June, 1907, or an increase of about 245 per cent. in three and one quarter years, which was followed within the next nine months by another sudden drop of $27,300,000.

These points of low earnings came within periods of general industrial depression which affected all lines of business to a greater or less degree. They are significant for the reason already given

and because they show the rather severe tests to which this great combination has been subjected during its brief career. The corporation has weathered the storms of these two periods without material damage and has shown its great powers of resistance.

In explaining the chart the author states that the main object of the analysis is to assist in forming an opinion as to the standing of the corporation's securities, and he concludes that the bonds rank with the highest grade industrial bonds and the better grades of railroad bonds. He regards the preferred stock as one of the most desirable investments of its kind in the industrial list and predicts that the comomn stock will establish a good dividend record.

On the whole the record graphically presented by means of the chart seems to warrant the author's conclusions and to indicate that the corporation has been well managed. It is apparent that the tremendous earnings of this great organization have justified the optimism of its promoters whether such earnings can be in themselves justified in the final analysis or not.

The following figures for the quarters ending December 31, 1909, and March 31, 1910, may be interesting to those who would like to continue the curves on the chart.

	Dec. 31.	Mar. 31.
Sinking Fund Sub. Cos.	.4	.3
Depreciation and Reserve Funds	6.1	5.8
Interest and Sinking Funds, U. S. S. Corp. Bonds	7.3	7.3
Preferred Stock Div. U. S. S. Corp.	6.3	6.3
Common Stock Div. U. S. S. Corp.	8.9	6.4
Approp. for Additional Property, etc.	5.0	5.0
Reserve Fund to cover Advanced Mining Royalties	3.2*	—
Surplus	3.8	6.5
Total Quarterly Earnings	41.0	37.6

*This is a new sub-division which evidently will not appear regularly.

C. W. D.

OCCUPATION MORTALITY STATISTICS OF NEW JERSEY.

The State Board of Health of New Jersey is making a laudable attempt to collect and compile mortality data by occupations. The classification includes 93 specific occupations the mortality data for which are separated by age and by principal causes of death. The age groups are 10–14, 15–19, 20–29, 30–39, 40–49, 50–59, 60–69, 70–79, 80–89 and over 90. The causes of death specified are consumption, pneumonia, cancer, suicide, diseases of the nervous system and of the organs of sense, diseases of the circulatory system, diseases of the respiratory system (consumption and pneumonia excepted), diseases

of the digestive system, diseases of the genito-urinary system and adnexa.

Unfortunately, the data are not presented by age for all other causes than those specified, nor are the data presented by age for the grand total of deaths in each occupation. It is, therefore, not possible to calculate the proportional mortality from any specified disease, with distinction of age, and as the number of employees in any given occupation is not known it is not possible to calculate death rates. In the 1908 returns, however, the deaths from "all other causes" and from "all causes" will be compiled by age and the occupation data will then be of great practical value. It is understood that the deficiencies in this respect for previous years will be remedied so far as possible by a review of the unclassified returns for the period 1901–1907 during which the mortality statistics have been compiled by occupation.

It is also very unfortunate that accidents have not been separately reported by occupation. Of what practical value are the mortality returns for such occupations as boatmen, electricians, firemen, fishermen, linemen, miners, railroad employees, sailors, etc., if the deaths from accident are not given? It is to be hoped that that defect will likewise be remedied in the future vital statistics of New Jersey, and that the missing data for the earlier years may be supplied in a summary table for a five-year period, at least. New Jersey will then be far ahead of any other American state in the quality and quantity of its mortality data by occupation. New Jersey is one of the foremost states in glass, hat, pottery and silk manufacture and holds an important position in other industrial pursuits which are of the greatest possible interest from the viewpoint of industrial hygiene. The State Board of Health of New Jersey is therefore to be congratulated upon having already made public so large a body of facts illustrative of the principal causes of death in 93 specific occupations, and when the defects in the data, here referred to, are remedied their annual reports will form an invaluable addition to our knowledge of occupation mortality.

F. S. C.

A REPORT ON THE TAXATION LOSS OF THE STATE OF RHODE ISLAND, 1910, BY JOINT SPECIAL COMMITTEE OF THE GENERAL ASSEMBLY OF THE STATE OF RHODE ISLAND.

This report contains the results of an extended investigation made by a committee consisting of Robert S. Franklin, William M. P. Bowen, Oscar A. Bennett, Zenas W. Bliss and William C. Bliss, in accordance with a special resolution passed by the General Assembly, May 6, 1909.

The report briefly sets forth the condition of the State laws in regard to taxation and recommends certain changes. In reaching its conclusions the committee was influenced by the experience of other states and the advice of a number of experts on taxation, and it is interesting to note that it has embodied in its recommendations a considerable number of advanced ideas in regard to taxation.

The report is especially valuable, not only because of the conclusions reached by the committee, but by reason of the statistical data which it contains and upon which the conclusions are based. These data are classified under several heads: Taxation of Inheritances, State Taxation of Corporations, Taxation on Mortgages and other Personal Property, and State Supervision of Local Assessment and Equalization of Taxes. Under these headings not only are the laws of many states summarized but the receipts for a period of years from these various kinds of taxes are given in tabular form. As one might naturally expect, however, the report contains fullest information in regard to the State of Rhode Island itself. It is interesting to note, among other things, that the valuation of personal property increased more rapidly in that state during the ten years from 1899 to 1908 than the valuation of real estate. For the State as a whole the valuation of real estate increased 21.3 per cent. during the decade and that of personal property 35.8 per cent. This is rather an unusual experience in view of the fact that the opportunities for concealing personal property are increasing and the tendency in that direction in other places is very marked. The large city of Providence shows an even greater proportionate increase in personal property; the increase in that city was 18.9 per cent. in real estate and 38.8 per cent. in personal property. In spite of this fairly favorable showing the committee is convinced that a departure should be made from the general property tax principle and that personal property, particularly that of an intangible kind, should be taxed at a special rate, namely 48 cents per hundred dollars of valuation, 39 cents of which should go to the city or town where assessed and 9 cents to the State.

The report strikes one as being progressive without being radical, and it is to be hoped that most of the recommendations made will be enacted into law. If they are the State of Rhode Island will have a system of taxation far in advance of most of the states of the Union.

C. W. D.

CORRECTION.

In the account of the twelfth biennial session of the International Statistical Institute, contained in the December (1909) number of these publications, mention should have been made of the fact that the vacancy caused by the death of the President, von Inama-Sternegg, was filled by the unanimous election of Luigi Bodio, for many years the efficient Secretary-General of the Institute and now Senator of the Kingdom of Italy. A sketch of his life and a list of his publications may be expected for the next issue of the publications.

W. F. W.

Publications of the American Statistical Association.

The New Series of the Association, issued quarterly, will be sent to members of the Association on payment of the annual dues of two dollars. The subscription price is also two dollars per annum. Separate copies can also be purchased upon remittance of the advertised prices.

The following list of titles includes only the most important papers published in the several issues. There is in addition a considerable amount of miscellaneous material.

VOL. I. 1888-1889. Nos. 1-8. Pp. 492.

As No. 4 is out of print, it is impossible to furnish this volume complete. Price without No. 4, $3.50.

No. 1. Statistics of Water Power Employed in Manufacturing in the United States. GEORGE F. SWAIN. 75 cents.

Nos. 2, 3. Park Areas and Open-Air Spaces in American Cities. E. R. L. GOULD.

Key to the Publications of the United States Census, 1790-1887; with Occasional References to Other Statistical Works. EDWARD CLARK LUNT. $1.00.

No. 4. Life Insurance in the United States. WALTER C. WRIGHT. (Out of print.)

No. 5. Notes on the Statistical Determination of the Causes of Poverty. AMOS G. WARNER. REMARKS. SAMUEL W. DIKE.

Statistics of Divorce in the United States and Europe. SAMUEL W. DIKE. 50 cents.

No. 6. American Railroad Statistics. ARTHUR T. HADLEY.

Statistics of Municipal Finance. HENRY B. GARDNER. 75 cents.

No. 7. Prison Statistics of the United States for 1888. ROLAND P. FALKNER. 50 cents.

No. 8. Finance Statistics of the American Commonwealths. EDWIN R. A. SELIGMAN.

Divorce in France. BENJAMIN F. KELLER.

Relief of the Poor in Germany. A. G. WARNER. $1.00

VOL. II. 1890-1891. Nos. 9-16. Pp. 470. PRICE $4.00.

No. 9. Mortgage Statistics. GEORGE K. HOLMES. 50 cents.

No. 10. The Study of Statistics in Italian Universities. CARROLL D. WRIGHT.

Statistics of Private Corporations. ROLAND P. FALKNER. 50 cents

Nos. 11, 12. Statistics of the Colored Race in the United States. FRANCIS A. WALKER.

How to Make Statistics Popular. CHARLES F. PIDGIN.

Retail Prices in Boston and Vicinity. WALTER F. COOK. $1.00.

No. 13. The Growth of Cities in Massachusetts. HORACE G. WADLIN.

Rate of Natural Increase of Population in United States. HERMAN HOLLERITH.

The First Census of Massachusetts. SAMUEL A. GREEN.

The Commercial Death Rate. ALBERT C. STEVENS. $1.00.